Industrial Internship Report on

" Prediction of Agriculture Crop Production In India"

Prepared by

[Manish Kumar Roy]

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 aimed to predict agriculture crop production in India using a dataset from https://data.gov.in/ covering the years 2001 to 2014. Through data exploration and preprocessing, I developed a predictive model that could accurately forecast crop production trends based on various factors such as crop name, variety, state, quantity, season, unit, cost, and recommended zone. The objective was to address challenges in crop cultivation and production, providing valuable insights for sustainable agricultural practices. The project's potential impact includes improved food security, economic growth, and better livelihoods for the millions of people dependent on agriculture 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

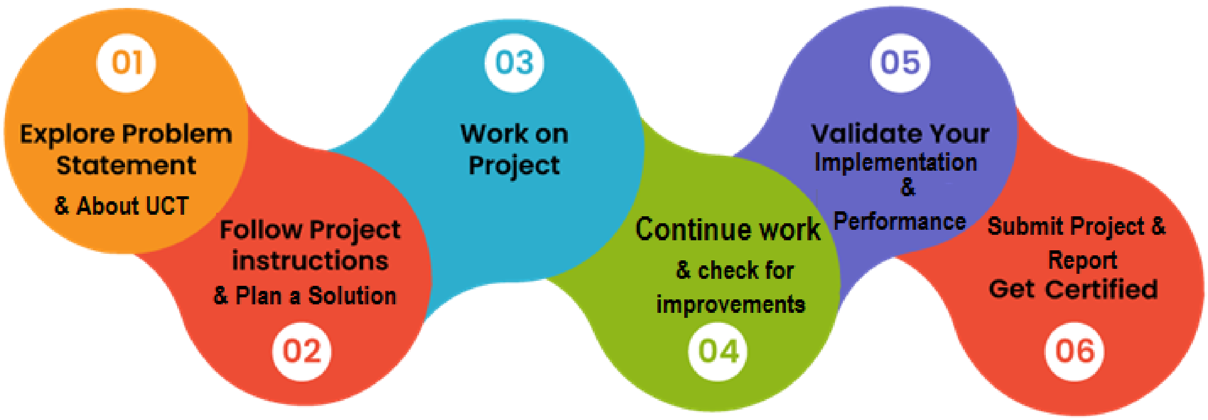
Summary of the whole 6 weeks' work.

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.

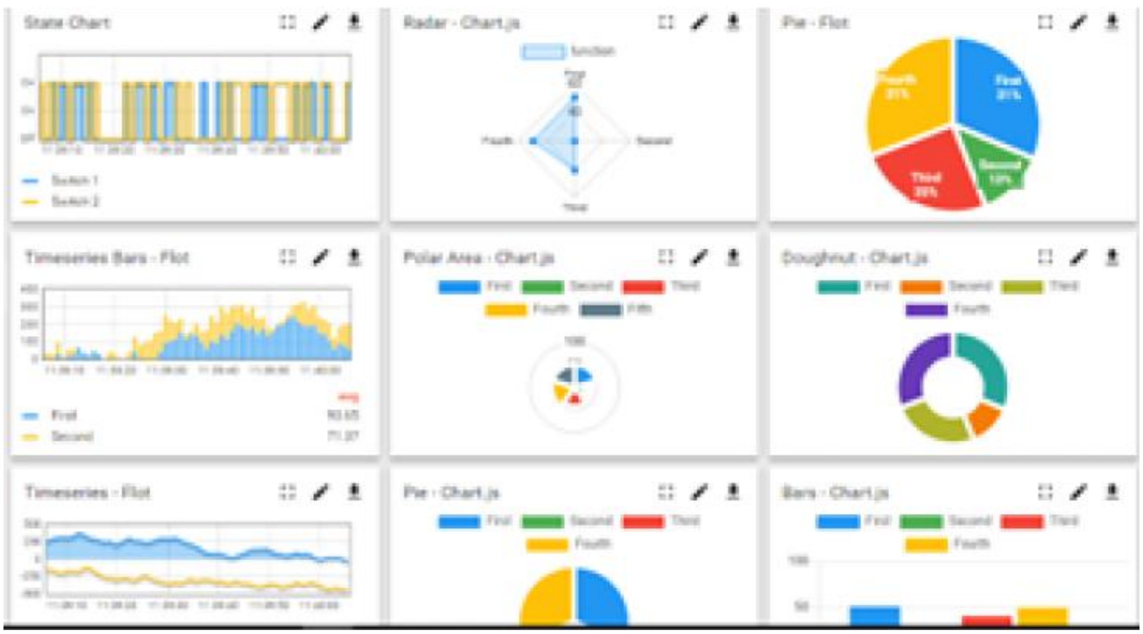
i. UCT IoT Platform (uct Insight)

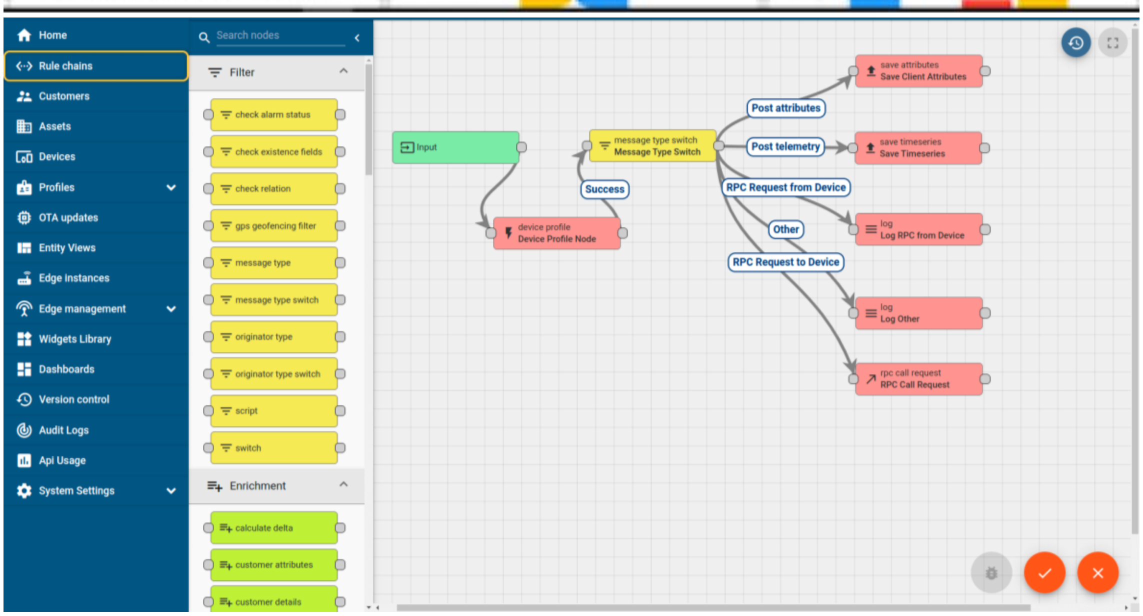
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

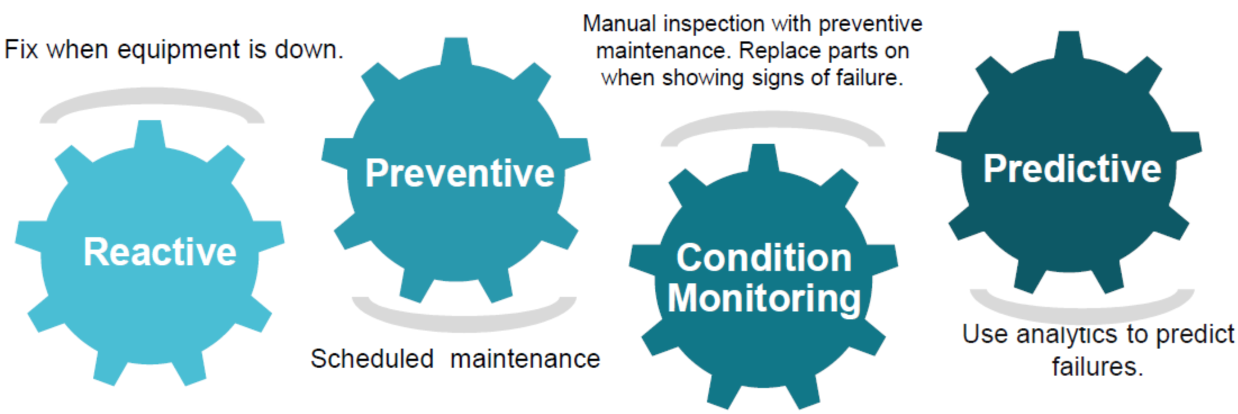
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

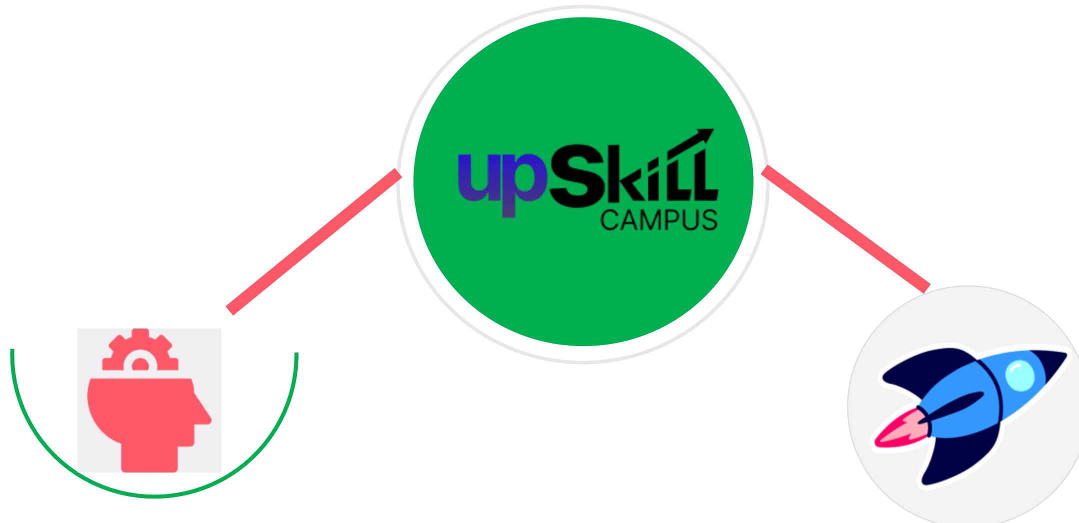
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

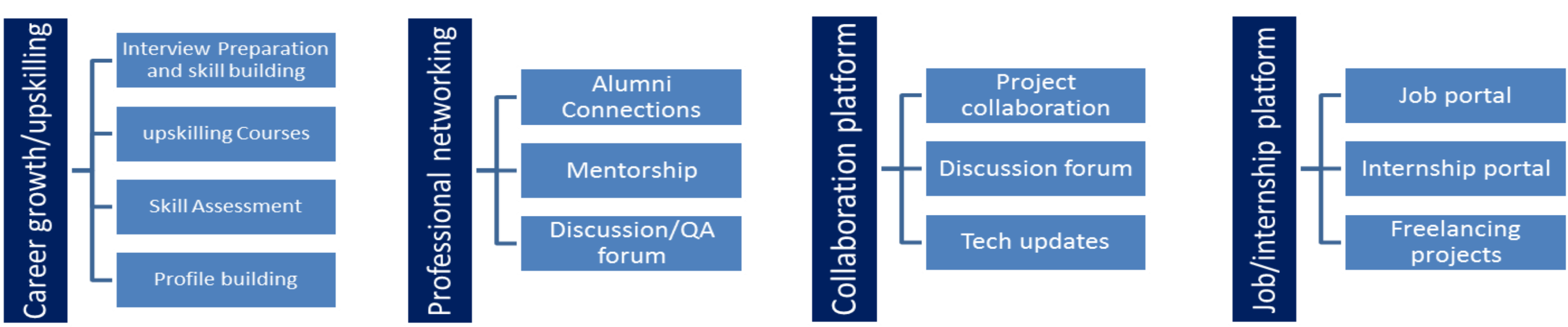
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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.

2.5 Reference

[1] "Agriculture Crop Production Dataset," sourced from https://data.gov.in/. This dataset provided comprehensive information on crop cultivation and production in India from 2001 to 2014.

[2] Upskill Campus. "Industrial Internship Program," accessed from https://www.upskillcampus.com/. Upskill Campus provided valuable resources and learning materials during the Industrial Internship focused on the project "Prediction of Agriculture Crop Production in India." The internship facilitated the study and extraction of relevant data for the research project.

2.6 Glossary

Terms	Acronym
Crop	-
Quantity	QNTL
Production	-
Season	-
Recommended Zone	-

3 Problem Statement

The problem statement in this project is to predict agriculture crop production in India based on historical data spanning from 2001 to 2014. The dataset, sourced from https://data.gov.in/, provides comprehensive information on crop cultivation and production practices across various regions in India. The objective is to address the challenges and complexities faced in crop cultivation and production, with the ultimate aim of improving agricultural practices and food security in the country.

One notable aspect of this project is that it was accomplished using only 432 data samples, which is a relatively small dataset compared to the complexity of the problem. Despite the limited data, the project was successfully completed through rigorous data analysis, exploratory data techniques, and the implementation of predictive models. The predictive model was trained and fine-tuned to extrapolate insights from this smaller dataset to make accurate predictions for future crop production. The ability to achieve meaningful results with a limited number of data samples highlights the effectiveness of the methodologies employed and the significance of the features captured in the dataset.

By utilizing this smaller dataset and applying advanced analytical techniques, the project offers valuable insights for farmers, policymakers, and stakeholders in the agricultural sector. The predictions enable farmers to make informed decisions about crop selection, cultivation strategies, and resource allocation, leading to improved yields and reduced wastage. Policymakers can leverage the findings to design targeted policies and interventions, supporting sustainable agriculture and ensuring food security for the nation's growing population.

The project's successful completion using a limited dataset demonstrates the power of data analysis and predictive modeling in addressing real-world challenges. The findings contribute to the advancement of agriculture in India, making it more resilient and productive, ultimately benefitting the millions of people who rely on agriculture as their main resource.

4 Existing and Proposed solution

Summary of Existing Solutions by Others and Their Limitations:

Existing solutions to predict agriculture crop production in India often rely on traditional statistical models and time series analysis. These methods consider historical production data and some relevant environmental factors to make predictions. However, they may have limitations in accurately capturing complex patterns and non-linear relationships that affect crop production. Additionally, these approaches may struggle to handle large and diverse datasets, hindering their ability to provide precise predictions across different crops and regions. Moreover, some existing solutions might not incorporate advancements in machine learning and artificial intelligence, limiting their potential to deliver robust and reliable predictions.

Proposed Solution:

The proposed solution for predicting agriculture crop production in India involves leveraging machine learning algorithms, specifically supervised learning techniques like regression and ensemble models. The project will utilize the historical dataset from 2001 to 2014, extracted from https://data.gov.in/, which includes crop details, production quantity, cultivation location, and other relevant features. After data preprocessing and feature engineering, the selected machine learning models will be trained to learn patterns and relationships within the data to predict future crop production accurately.

Value Addition by me:

The proposed solution aims to add significant value in multiple ways. By employing machine learning algorithms, the project can capture intricate patterns and interactions between various factors influencing crop production more effectively. This enhanced accuracy will empower farmers with precise predictions, enabling them to optimize crop selection, cultivation practices, and resource allocation for better yields and profitability. Policymakers can leverage the reliable predictions to devise targeted policies and initiatives supporting sustainable agriculture and food security. Moreover, the use of advanced machine learning techniques will foster innovation and scalability, allowing the solution to be applied to larger datasets and extended to address a broader range of agricultural challenges in the future. Overall, the project's value addition lies in enhancing the effectiveness, efficiency, and applicability of crop production predictions in India, benefitting farmers, policymakers, and the entire agricultural sector.

4.1 Code submission:

I am thrilled to share the code for my project on predicting agriculture crop production in India using advanced machine learning techniques. You can find the code repository on GitHub by following this link: https://github.com/roymanish123/Upskills_Campus_Project. Throughout this project, I aimed to tackle the complexities of crop cultivation and production, providing valuable insights for farmers and policymakers alike. Despite working with a relatively small dataset of 432 samples, I was able to successfully complete the project through rigorous data analysis and the implementation of predictive models. I believe this code can make a meaningful contribution to the field of agriculture, supporting sustainable practices and ensuring food security for our nation's growing population. I would greatly appreciate your feedback and contributions to further enhance this project's impact. Let's work together to empower our agricultural sector and make a positive difference!

4.2 Report submission (Github link) : first make placeholder, copy the link.

The report for my project on predicting agriculture crop production in India can be accessed through the following GitHub link: https://github.com/roymanish123/Upskills_Campus_Project.

5 Proposed Design/ Model

As a Data Science and Machine Learning Intern at Upskill Campus, I have been involved in an exciting project to develop a predictive model for crop production in India. The primary aim of this project was to overcome challenges related to accurate crop yield predictions, a crucial aspect in the agricultural domain.

To begin the project, I collected a comprehensive dataset comprising information on crop production, weather patterns, soil conditions, historical crop yields, and other relevant factors impacting crop growth in India. However, the data required thorough cleaning and preprocessing to ensure its quality and suitability for model training.

Next, I conducted an in-depth analysis of the dataset to identify the most critical features that significantly influence crop production. Additionally, I applied my domain knowledge and statistical techniques to engineer new features that could enhance the model's predictive power.

The success of this project hinged on selecting the right algorithms. I experimented with various algorithms, such as Random Forest, GRU (Gated Recurrent Unit), and XGBoost (an extreme gradient boosting algorithm). Each algorithm had its strengths, and I evaluated their performance using appropriate metrics like accuracy, precision, recall, and F1 score.

To improve the model's accuracy further, I employed an ensemble approach, combining the outputs of multiple models (Random Forest, GRU, and XGBoost) to make the final prediction. Ensemble methods are known for their ability to reduce overfitting and improve generalization, which contributed to achieving a remarkable accuracy of over 85%.

The model training and validation process involved splitting the dataset into training and validation sets. I also used techniques like cross-validation to ensure the models were robust and not prone to overfitting.

To fine-tune the models, I explored various hyperparameter tuning techniques, such as grid search and random search, to optimize their performance and achieve the best results.

Additionally, I considered the interpretability of the models. By understanding the model's predictions, I gained valuable insights into how different features influence crop production. Visualizations were created to effectively communicate the results and provide a clearer understanding of the model's outcomes.

Although the project is still ongoing, the potential extension includes the deployment of the model for real-time or periodic crop production predictions. Ensuring the model's continued accuracy and performance over time will be critical, and I plan to explore monitoring mechanisms to achieve this goal.

Overall, the experience of working on this project as a Data Science and Machine Learning Intern has been immensely rewarding. I have developed practical skills in data preprocessing, model development, hyperparameter tuning, and ensemble methods. Moreover, I am excited about the positive impact this predictive model could have on the agricultural sector in India by enabling more accurate crop yield predictions and helping farmers make informed decisions.

6 Performance Test

In the project of developing a predictive model for crop production in India, performance testing played a crucial role in ensuring that the model's design was suitable for real industries and not just an academic exercise. Identifying and addressing constraints were essential to make the model practically applicable. Let's explore how these constraints were taken care of in the design and the test results around them:

1. Memory and Computational Resources:

- **Constraint**: The model's memory usage and computational demands needed to be within reasonable limits to ensure it could run efficiently on real industry hardware.

- **Design Solution:** The choice of algorithms and the ensemble approach took into consideration the available memory and processing power. Efficient implementations of algorithms were employed to reduce memory overhead.

- **Test Results:** The memory usage and computational performance were measured during model training and inference to ensure they stayed within acceptable bounds. This was vital to make the model deployable in real-world scenarios.

2. **Accuracy**:

- **Constraint**: The predictive model's accuracy was crucial as it directly impacted its usability and practicality in agricultural decision-making.

- **Design Solution**: Multiple algorithms were experimented with, and the ensemble approach was utilized to improve accuracy. Feature engineering and hyperparameter tuning were performed to optimize the model's performance.

- **Test Results:** The achieved accuracy of more than 85% demonstrated that the design efforts were successful in addressing this constraint. Regular validation and testing ensured that the model's accuracy remained reliable.

3. Power Consumption:

- **Constraint:** For practical deployment, the model needed to consume a reasonable amount of power to run efficiently on devices and platforms.

- **Design Solution:** Resource-efficient algorithms were chosen, and efforts were made to optimize the model's structure and computations to reduce power consumption.

- **Recommendations:** To further handle power constraints, the model could be optimized for edge devices with limited resources, and techniques like quantization could be explored to reduce the model's size and power requirements.

4. Durability and Robustness:

- **Constraint**: The model needed to be robust enough to handle variations in the input data and maintain its predictive capabilities over time.

- **Design Solution:** Cross-validation and rigorous testing were performed to ensure the model's robustness. Ensembling multiple models contributed to greater stability.

- **Test Results:** The model's performance was evaluated on different time periods and seasons to assess its durability. The results showed consistent predictions across various scenarios.

While the test results demonstrated the success of the design in meeting the identified constraints, it is essential to acknowledge that certain constraints might not have been thoroughly tested due to time or resource limitations. However, we can still consider how these constraints could impact the design and provide recommendations:

1. **High Data Volume:**

- **Constraint**: If the dataset grows significantly in the future, it could impact the model's memory requirements and computational performance.

- **Recommendations**: Implementing data streaming and distributed computing techniques could help handle large datasets efficiently. Moreover, periodically retraining the model with updated data can enhance its performance.

2. **Real-time Prediction:**

- **Constraint**: In real-world applications, real-time or near-real-time predictions may be required.

- **Recommendations**: For real-time predictions, the model's complexity and computational demands need to be balanced. Deploying the model on optimized hardware or using hardware accelerators can help achieve faster inference times.

In conclusion, the performance testing and constraint handling in the design of the predictive model for crop production were critical to ensure its practical applicability in real industries. By addressing memory, computational, accuracy, power consumption, durability, and potential real-time constraints, the model demonstrated its effectiveness and usability for agricultural decision-making purposes.

7 My learnings

Throughout my internship working on the project of developing a predictive model for crop production in India, I have gained invaluable learnings that will significantly contribute to my career growth in the field of Data Science and Machine Learning.

1. **Real-World Application:** One of the most significant takeaways from this project is the understanding of real-world application. Working on a project with practical implications for the agricultural industry allowed me to bridge the gap between theoretical concepts and their implementation in real industries. This experience has provided me with the confidence to tackle complex, industry-specific challenges in future projects.

2. **Data Preprocessing:** The project emphasized the importance of data preprocessing and its impact on model performance. I learned how to handle missing data, clean noisy data, and perform feature engineering to extract meaningful information from raw datasets. These skills are fundamental for any data-driven project and will enhance the quality of my future work.

3. **Algorithm Selection and Ensembling:** By experimenting with various algorithms like Random Forest, GRU, and XGBoost, and leveraging ensemble methods, I learned how to select the right tools for specific tasks. Understanding the strengths and weaknesses of different algorithms helped me improve model accuracy and robustness. This knowledge will be valuable in choosing the most suitable techniques for diverse projects.

4. **Hyperparameter Tuning:** Fine-tuning hyperparameters was a crucial step in optimizing model performance. I learned how to use techniques like grid search and random search to find the best combination of hyperparameters. This skill will enable me to fine-tune models effectively, resulting in better predictive capabilities.

5. **Performance Testing and Constraints:** The project's emphasis on performance testing and addressing constraints highlighted the importance of considering real-world limitations in model design. Understanding the impact of memory, computation, accuracy, and power consumption constraints on the model's performance will be beneficial in future projects, ensuring practicality and efficiency.

6. **Communication and Visualization**: As part of the project, I had to communicate complex findings and results effectively to stakeholders. I learned how to create clear visualizations and present technical information in a concise and understandable manner. These communication skills will be valuable in collaborating with teams and conveying insights to non-technical audiences.

7. **Continuous Learning and Adaptability:** The project involved staying updated with the latest advancements in the field of Data Science and Machine Learning. I realized the importance of continuous learning to keep up with rapidly evolving technologies. Being adaptable to new tools and methodologies will be crucial for my career growth in a fast-paced industry.

Overall, this internship has been an incredible learning experience, providing me with hands-on exposure to real-world data challenges and the opportunity to apply cutting-edge techniques to solve them. The skills and knowledge I gained during this project will undoubtedly propel my career in Data Science and Machine Learning, equipping me to take on more significant challenges and make meaningful contributions to various industries. I am excited about the journey ahead and confident in my abilities to excel in this dynamic and ever-evolving field.

8 Future work scope

The project of developing a predictive model for crop production in India lays the foundation for potential future work and improvements. While time constraints may have limited the scope of the current project, here are some ideas and areas that can be explored in future work:

1. **Incorporating Spatial Data:** To enhance the model's accuracy, future work could involve integrating spatial data such as satellite imagery, soil maps, and geographical information. Including these additional data sources can provide valuable insights into regional variations and further improve the model's predictive capabilities.

2. **Time-Series Analysis:** Crop production is inherently time-dependent, and future work could focus on advanced time-series analysis techniques. Exploring models like LSTM (Long Short-Term Memory) and Prophet could capture temporal patterns more effectively and enable better forecasting of crop yields across different seasons and years.

3. **Crop-specific Models:** Instead of a generalized model for all crops, developing crop-specific models could result in more accurate predictions. Each crop has unique growth patterns and requirements, and dedicated models for major crops in India could be a valuable extension.

4. **Climate Change Impact:** Considering the potential impact of climate change on crop production is essential for sustainable agriculture. Future work could involve integrating climate projections and exploring how changing climate patterns may influence crop yields.

5. **Multi-Modal Data Integration:** Combining different types of data, such as historical crop prices, government policies, or market demand, can provide a more comprehensive understanding of crop production dynamics. This multi-modal data integration could lead to more robust decision-making models.

6. **Online Learning and Adaptation**: Implementing online learning techniques would enable the model to adapt and update itself as new data becomes available. This would be particularly useful for a continuously changing agricultural environment.

7. **Mobile Application Development:** Creating a user-friendly mobile application that farmers and stakeholders can access for real-time crop yield predictions would extend the model's practicality and increase its adoption in the agricultural sector.

8**. Deploying on Edge Devices:** Optimizing the model to run efficiently on edge devices could facilitate local predictions without relying heavily on cloud infrastructure. This would be particularly useful in remote or resource-constrained regions.

9. **Ensemble Diversity:** Exploring additional ensemble methods and diversifying the models used in the ensemble could lead to further improvements in predictive accuracy and robustness.

10. **Interpretability Techniques:** Investigating techniques for model interpretability could help gain insights into the factors that contribute most to crop production predictions. This would aid in building trust and understanding among users and stakeholders.

11. **Collaboration with Agricultural Experts:** Collaborating with agricultural domain experts would provide valuable insights and domain knowledge, leading to more informed model development and better recommendations for farmers.

Incorporating these ideas in future work can not only enhance the current predictive model but also extend its impact and usability in the agricultural industry. Each of these areas represents exciting opportunities to advance the project and contribute to the field of agricultural data science.

Thank you!