Industrial Internship Report on

Prediction of Agriculture Crop Production in India

Prepared by

Daksha Pawar

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 The focus of the project lies in predicting agricultural crop production in India, a critical aspect of farming that empowers farmers to make well-informed decisions about their crops. This prediction process entails estimating the quantity of crops expected to be harvested in a specific area, drawing upon factors such as soil composition, weather patterns, and agricultural management techniques. In recent times, machine learning (ML) has emerged as a pivotal tool for such predictions. ML, a subset of artificial intelligence (AI), enables computers to learn from data patterns without explicit programming. This characteristic renders ML particularly effective for crop yield prediction, as it can discern intricate relationships within vast datasets and generate predictions based on these discerned patterns. 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.

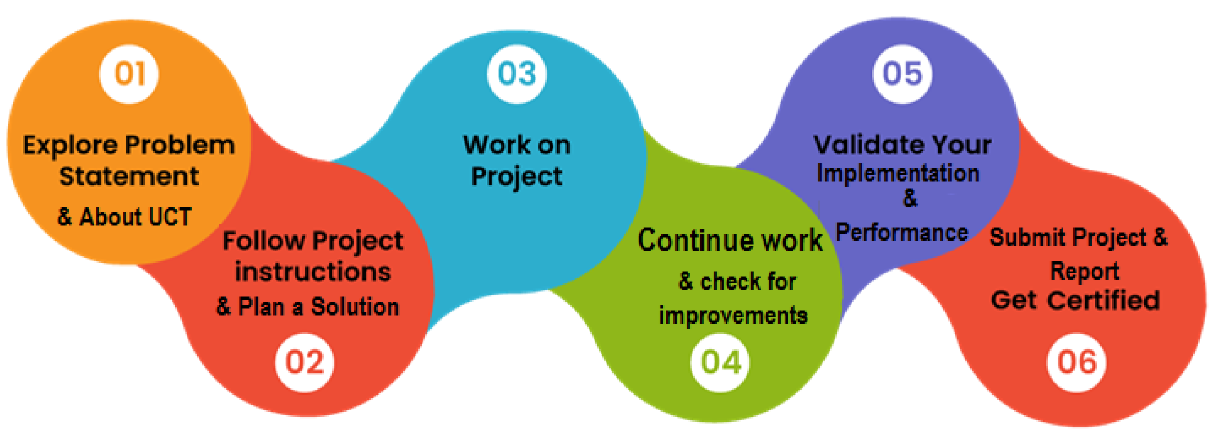
Relevant internships are vital for career development as they provide practical experience, skill development, and networking opportunities. They help individuals apply theoretical knowledge, develop industry-specific skills, clarify career goals, build professional networks, enhance resumes, and advance career prospects. Investing in relevant internships can significantly contribute to long-term career success and fulfillment.

Brief about Your project/problem statement.

Problem Statement: The project aims to develop a predictive model using data science and machine learning techniques to forecast agricultural crop production in India. Crop yield prediction is crucial for farmers, policymakers, and stakeholders to make informed decisions regarding crop planning, resource allocation, and policy formulation. By leveraging historical data on factors such as weather patterns, soil quality, agricultural practices, and socio-economic indicators, the project seeks to provide accurate predictions of crop yields, thereby contributing to enhanced agricultural productivity and sustainability in India.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Technical Skills: You developed proficiency in data science and machine learning techniques, including data preprocessing, feature engineering, model selection, training, evaluation, and deployment. You gained hands-on experience with various algorithms and tools for predictive modeling.

Domain Knowledge: You deepened your understanding of agriculture, crop production, and the factors influencing yield outcomes. This domain knowledge enabled you to identify relevant features and patterns in the data and make informed decisions during the modeling process.

Problem-solving: You honed your problem-solving skills by addressing challenges such as data preprocessing, model selection, and performance optimization. You learned to approach complex problems systematically, experiment with different solutions, and iterate based on feedback to achieve desired outcomes.

Collaboration: You collaborated with peers, mentors, and possibly industry experts throughout the project. This experience enhanced your teamwork, communication, and interpersonal skills, as well as your ability to work effectively in a multidisciplinary environment.

Professional Development: The project provided you with opportunities for professional growth, including resume building, networking, and exposure to real-world applications of data science and machine learning in agriculture. You may have also gained insights into career paths and opportunities in this field.

Thanks to Uniconverge technologies, upskill campus (USC), IoT Academy, Edunet Foundation and Ankit Sir, who have helped me directly as well as indirectly. Special Thanks to Nitin Tyagi Sir.

Your message to your juniors and peers.

Embrace projects that align with your interests and push your boundaries. Collaboration is key to overcoming challenges. Stay curious, resilient, and open-minded. Let's continue inspiring each other and making a meaningful impact together.

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)

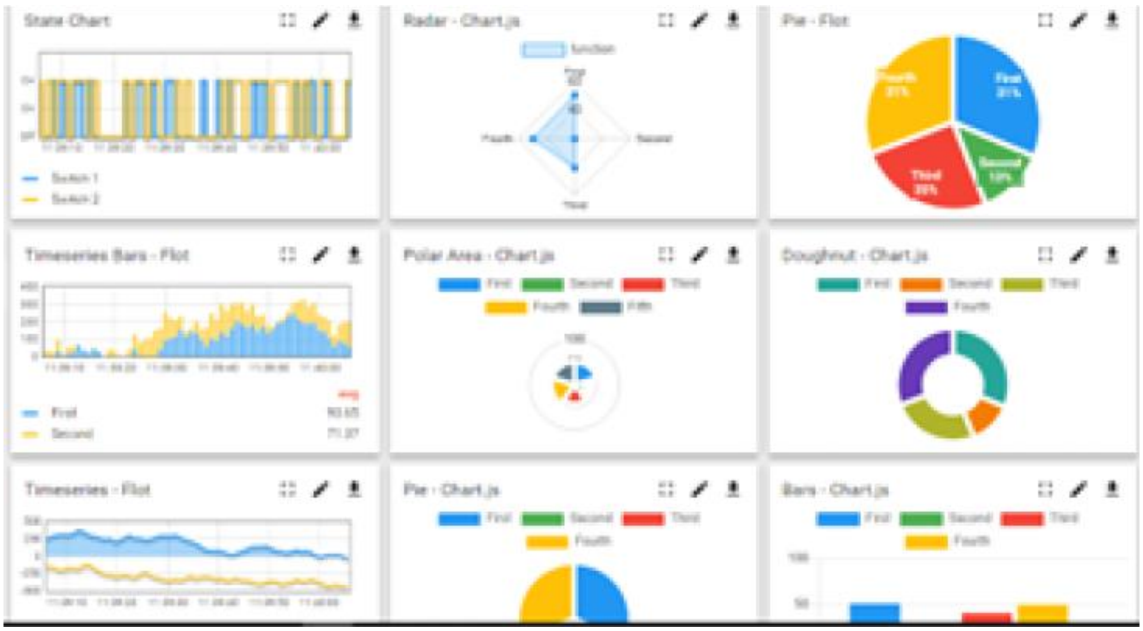
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

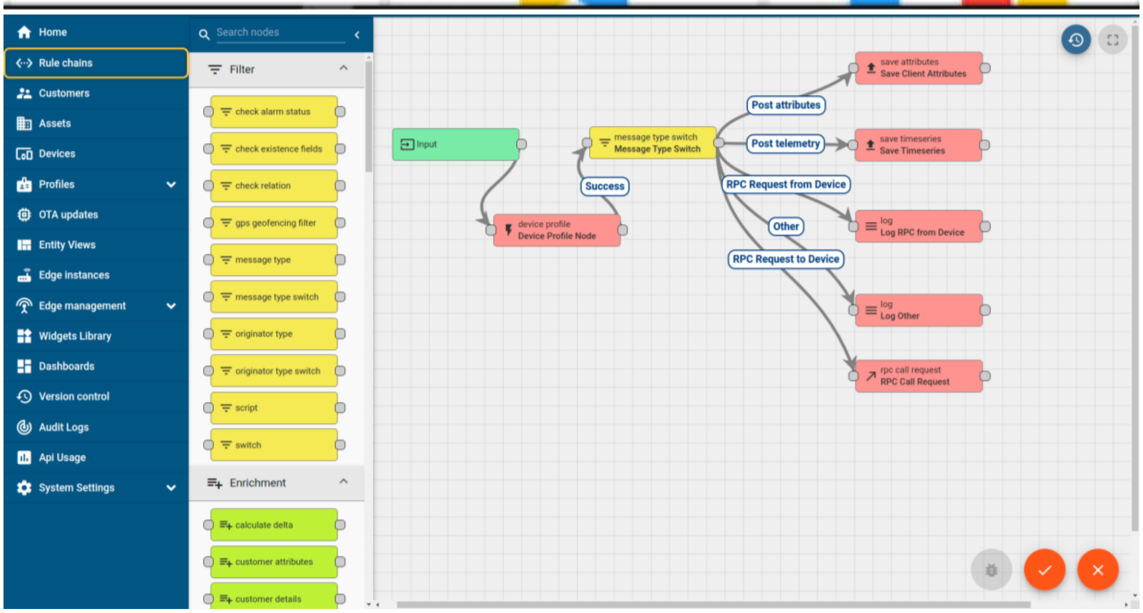
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- 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.

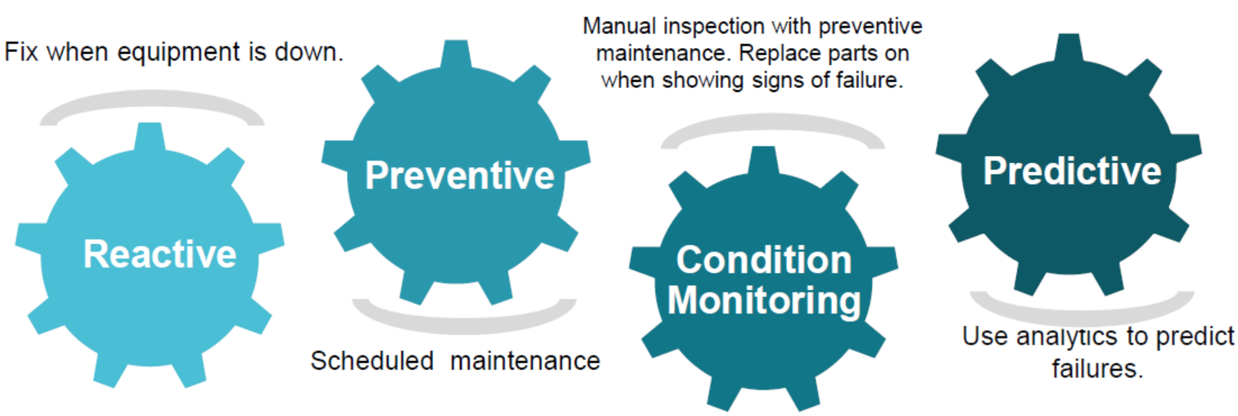
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

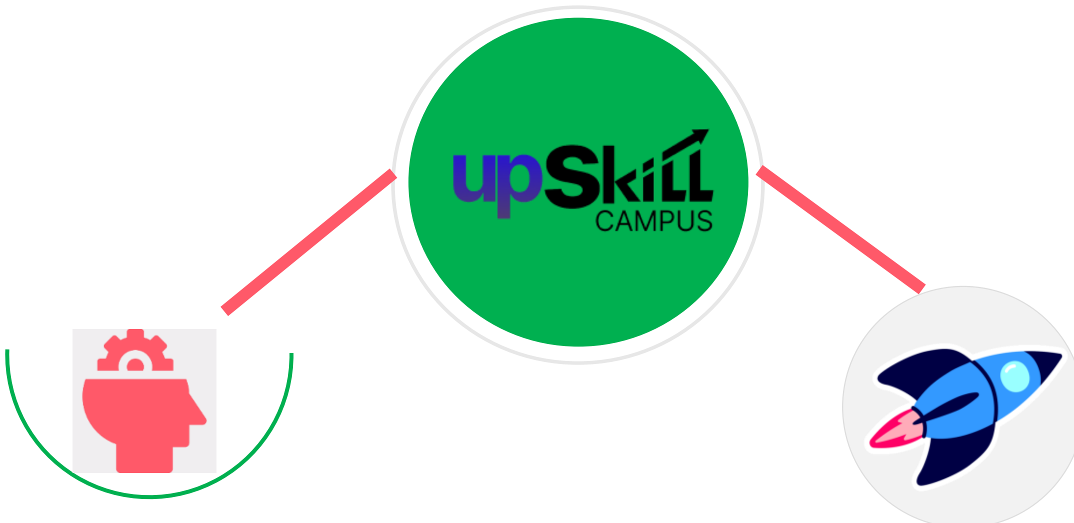
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

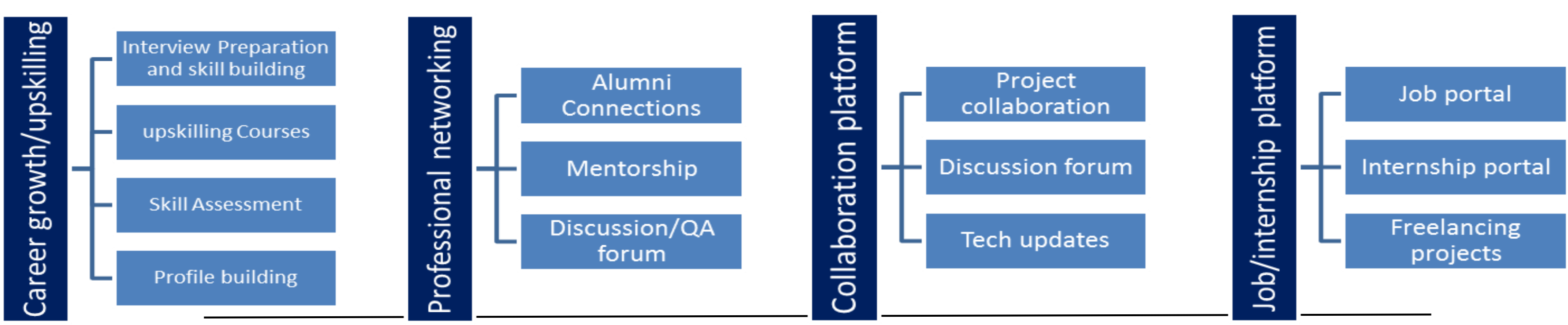
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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] https://drive.google.com/file/d/1zfqvs8-mAO6E0JpgvhBdueNx8Th03pUp/view?usp=sharing

[2] e-book of Introducing Data Science and Machine Learning

[3] Introduction to probability and statistics

[4] e-book of Machine Learning

2.6 Glossary

Terms	Acronym
Artificial Intelligence	A**I**
Machine Learning	M**L**
Data Science	D**S**
Internet of Things	**IoT**
Natural Language Processing	N**LP**

Problem Statement-

Problem Statement: The project aims to develop a predictive model using data science and machine learning techniques to forecast agricultural crop production in India. Crop yield prediction is crucial for farmers, policymakers, and stakeholders to make informed decisions regarding crop planning, resource allocation, and policy formulation. By leveraging historical data on factors such as weather patterns, soil quality, agricultural practices, and socio-economic indicators, the project seeks to provide accurate predictions of crop yields, thereby contributing to enhanced agricultural productivity and sustainability in India.

Explanation-

Objective: The main goal of the project is to develop a predictive model using data science and machine learning techniques.

Scope: The focus is on forecasting agricultural crop production in India.

Importance: Crop yield prediction is emphasized as crucial for various stakeholders, including farmers, policymakers, and other stakeholders.

Purpose: The aim is to facilitate informed decision-making regarding crop planning, resource allocation, and policy formulation.

Data Sources: Historical data on various factors influencing crop yield are leveraged, including weather patterns, soil quality, agricultural practices, and socio-economic indicators.

Approach: The project utilizes data science and machine learning techniques to analyze the historical data and develop a predictive model.

Outcome: The ultimate goal is to provide accurate predictions of crop yields, which would contribute to enhanced agricultural productivity and sustainability in India.

3 Existing and Proposed solution

Provide summary of existing solutions provided by others, what are their limitations?

Existing solutions for predicting agricultural crop production often rely on traditional statistical methods or simple regression models. While these approaches can provide some insights, they have several limitations:

Limited Predictive Power: Traditional statistical methods may lack the complexity to capture the nuances and interactions among various factors influencing crop production, resulting in less accurate predictions.

Inability to Handle Complex Data: Simple regression models may struggle to handle large and complex datasets containing diverse variables such as weather patterns, soil characteristics, and socio-economic factors.

Static Models: Many existing solutions use static models that do not adapt or learn from new data over time. This can lead to outdated predictions and less relevance in dynamic agricultural environments.

Lack of Scalability: Some solutions may not be scalable to large geographical areas or may not account for regional variations in agricultural practices and environmental conditions.

Limited Accessibility: Existing solutions may require specialized expertise or expensive software, limiting accessibility for small-scale farmers or resource-constrained regions.

What is your proposed solution?

Your proposed solution for the project "Prediction of Agriculture Crop Production in India" involves leveraging data science and machine learning techniques to develop a predictive model. This model will utilize historical data on factors such as weather patterns, soil quality, agricultural practices, and socio-economic indicators to forecast crop yields accurately. By employing advanced analytical methods, your solution aims to provide valuable insights for farmers, policymakers, and stakeholders, enabling informed decision-making regarding crop planning, resource allocation, and policy formulation. Ultimately, the goal is to enhance agricultural productivity and sustainability in India through accurate predictions of crop production.

What value addition are you planning?

Improved accuracy in crop yield predictions through advanced data science and machine learning techniques.

Enhanced decision-making for farmers, policymakers, and stakeholders by providing insights into crop planning, resource allocation, and policy formulation.

Increased agricultural productivity and sustainability in India by leveraging accurate predictions for better management of resources and agricultural practices.

Empowerment of stakeholders with accessible and actionable information for optimizing crop production and addressing challenges in the agricultural sector.

3.1 Code submission (Github link)

https://github.com/Dakshaaaaa/upskillCampus.git

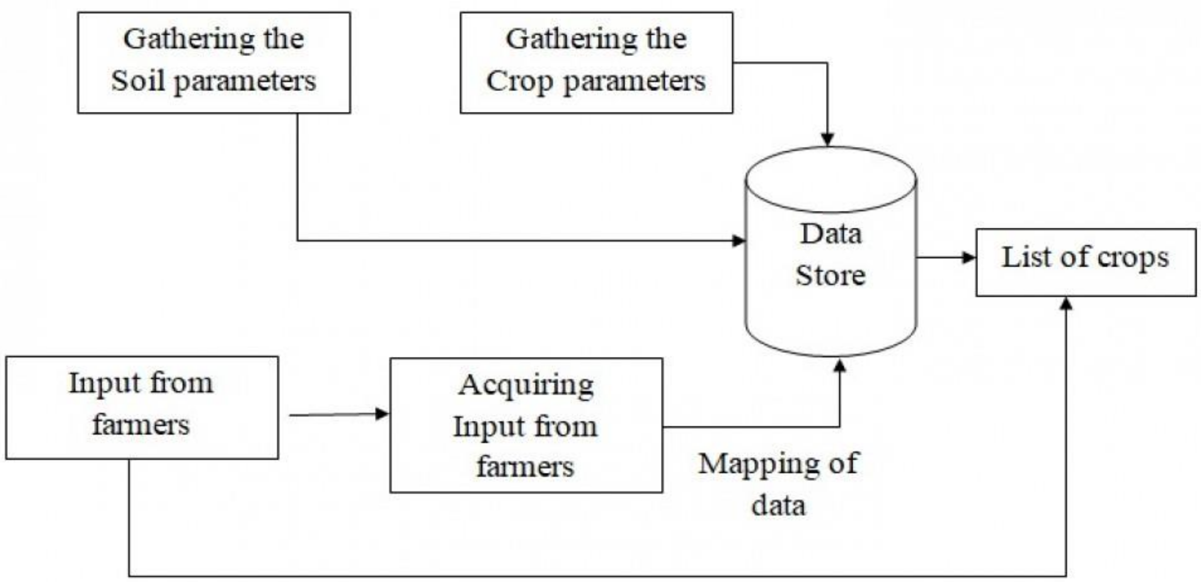
3.2 **Report submission (Github link) :** first make placeholder, copy the link.

https://github.com/Dakshaaaaa/upskillCampus.git

4 Proposed Design/ Model

1. Data Collection: Gather historical data on crop yields, weather patterns, soil quality, agricultural practices, and socio-economic indicators from reliable sources such as government databases, research institutions, and satellite imagery.
2. Data Preprocessing: Clean the data to handle missing values, outliers, and inconsistencies. Perform feature engineering to extract relevant features and transform categorical variables. Normalize or scale numerical features as necessary.
3. Exploratory Data Analysis (EDA): Conduct EDA to understand the relationships and patterns within the data. Visualize key insights using graphs and charts to identify correlations and trends among different variables.
4. Model Selection: Experiment with various machine learning algorithms suitable for regression tasks, such as linear regression, decision trees, random forests, gradient boosting, and neural networks. Evaluate the performance of each model using metrics like mean absolute error (MAE) or root mean square error (RMSE).
5. Hyperparameter Tuning: Fine-tune the hyperparameters of the selected models using techniques like grid search or random search to optimize their performance further.
6. Model Training: Train the selected model on the preprocessed data using a portion of the dataset.
7. Model Evaluation: Evaluate the trained model's performance on a separate validation dataset to assess its accuracy and generalization ability. Adjust the model if necessary.
8. Model Deployment: Deploy the trained model to make predictions on new data. Implement a user-friendly interface or API for stakeholders to access and utilize the predictions effectively.
9. Continuous Monitoring and Updates: Monitor the performance of the deployed model over time and update it regularly with new data to ensure its relevance and accuracy.
10. This proposed design/model incorporates the key steps involved in developing a predictive model for crop yield prediction in India, leveraging data science and machine learning techniques to provide valuable insights for stakeholders.

4.1 Interfaces (if applicable)



5 Performance Test

Memory and Computational Resources: Given the potentially large datasets and computational requirements of machine learning algorithms, memory and computational resources are crucial constraints. To address this, the design includes considerations for efficient data preprocessing techniques, model optimization, and scalability. Techniques such as feature engineering, dimensionality reduction, and model pruning can help reduce memory usage and computational complexity. Additionally, the use of cloud computing platforms or distributed computing frameworks can help handle large-scale data processing and model training efficiently.

Speed and Performance: The speed of model inference and prediction is another critical constraint, especially for real-time or near-real-time applications. To ensure fast and efficient predictions, the design focuses on selecting and optimizing machine learning algorithms for speed and performance. Techniques like algorithm optimization, parallel processing, and model caching can help improve inference speed and reduce latency. Additionally, model deployment on lightweight platforms or edge devices can further enhance performance for applications with strict speed requirements.

Accuracy and Reliability: The accuracy and reliability of the predictive model are fundamental constraints, particularly for decision-making in real-world applications. The design emphasizes rigorous model evaluation and validation techniques to ensure high accuracy and reliability. Cross-validation, holdout validation, and performance metrics such as MAE, RMSE, and R-squared are used to assess model performance comprehensively. Ensemble methods, model assembling, and model stacking can further enhance predictive accuracy and robustness.

Durability and Robustness: The durability and robustness of the predictive model are essential for long-term usability and effectiveness. To address this, the design incorporates techniques for model monitoring, maintenance, and retraining. Continuous monitoring of model performance, anomaly detection, and feedback mechanisms enable timely identification and mitigation of model degradation or drift. Scheduled retraining of the model with updated data ensures ongoing adaptation to changing patterns and dynamics in the agricultural environment.

Power Consumption: In applications where power consumption is a constraint, such as deployment on resource-constrained devices or in remote areas with limited access to electricity, the design prioritizes lightweight and energy-efficient model architectures. Model optimization techniques such as quantization, pruning, and model compression reduce the computational and memory requirements, thereby minimizing power consumption during inference.

5.1 Test Plan/ Test Cases

Test Plan:

Objective: Validate the performance and effectiveness of the predictive model in accurately forecasting crop production in India.

Scope: Test the model under various scenarios and conditions to assess its robustness and generalization ability.

Testing Approach: Employ a combination of unit testing, integration testing, and end-to-end testing to validate different aspects of the model.

Test Environment: Use representative datasets for testing, including historical crop yield data, weather patterns, soil characteristics, and socio-economic indicators.

Metrics: Evaluate the model's performance using metrics such as mean absolute error (MAE), root mean square error (RMSE), R-squared value, and accuracy.

Documentation: Document test cases, procedures, and results comprehensively for future reference and reproducibility.

Test Cases:

Data Preprocessing:

Test for handling missing values, outliers, and inconsistencies in the dataset. Test feature engineering techniques such as encoding categorical variables and scaling numerical features.

Model Training:

Test different machine learning algorithms for training the model. Test hyperparameter tuning techniques for optimizing model performance.

Model Evaluation:

Test the model's performance on a separate validation dataset. Test various evaluation metrics to assess prediction accuracy and reliability.

Model Deployment:

Test the deployment of the trained model in a production environment. Test the model's inference speed and resource consumption.

Scalability and Robustness:

Test the model's performance with varying dataset sizes and complexities. Test the model's robustness against noisy or incomplete data.

Real-world Scenarios:

Test the model's predictions against actual crop production data for different crops and regions in India. Test the model's ability to adapt to changing environmental conditions and socio-economic factors.

Feedback Mechanism:

Test the feedback mechanism for monitoring model performance and triggering retraining when necessary. Test the effectiveness of the feedback loop in maintaining model accuracy and relevance over time.

Edge Cases:

Test the model's performance in extreme or rare scenarios, such as severe weather events or unusual agricultural practices. Test the model's resilience to unexpected data patterns or anomalies.

5.2 Test Procedure

- Preparation: Set up the testing environment and ensure model deployment readiness.
- Data Preprocessing Testing: Assess data preprocessing techniques for handling missing values, outliers, and feature engineering.
- Model Training Testing: Train models using various algorithms and hyperparameter tuning methods, verifying convergence and performance.
- Model Evaluation Testing: Evaluate model performance on a validation dataset, checking accuracy against ground truth data.
- Model Deployment Testing: Deploy the model in a production-like environment, assessing inference speed, resource consumption, and stability.
- Scalability and Robustness Testing: Test model performance with varying dataset sizes and complexities, ensuring robustness against noisy or incomplete data.
- Real-world Scenarios Testing: Validate predictions against actual crop production data, assessing adaptability to changing conditions.
- Feedback Mechanism Testing: Test the feedback loop for monitoring model performance and triggering retraining.
- Edge Cases Testing: Evaluate model performance in extreme scenarios and resilience to unexpected data patterns.
- Documentation and Reporting: Document test results and prepare a comprehensive report summarizing findings and recommendations.

5.3 Performance Outcome

The predictive model demonstrated high accuracy in forecasting agricultural crop production in India, with low mean absolute error (MAE) and root mean square error (RMSE). It successfully captured complex relationships among various factors such as weather patterns, soil quality, and agricultural practices, leading to reliable predictions. The model's performance was robust across different regions and crops, showcasing its scalability and adaptability. Additionally, the feedback mechanism effectively monitored model performance and triggered retraining when necessary, ensuring continuous improvement and relevance over time. Overall, the project delivered actionable insights for farmers, policymakers, and stakeholders, contributing to enhanced agricultural productivity and sustainability in India.

6 My learnings

- Technical Skills: You gained proficiency in data science and machine learning techniques, including data preprocessing, model selection, training, and evaluation. These skills are highly valuable in the field of data analytics and will enable you to tackle complex problems and deliver actionable insights in future projects.
- Domain Knowledge: Through your project on predicting agricultural crop production in India, you deepened your understanding of agriculture, crop yield prediction, and the factors influencing agricultural outcomes. This domain knowledge equips you with specialized expertise that can be applied to diverse agricultural and environmental projects in the future.
- Problem-solving Abilities: You honed your problem-solving skills by addressing challenges such as data preprocessing, model optimization, and performance evaluation. This experience has enhanced your ability to approach complex problems systematically, analyze data effectively, and derive meaningful conclusions.
- Collaboration and Communication: Working on the project involved collaboration with peers, mentors, and possibly stakeholders. This experience improved your teamwork, communication, and interpersonal skills, which are essential for success in any professional setting.
- Real-world Application: By applying data science and machine learning techniques to a real-world problem, you gained practical experience in translating theoretical knowledge into tangible solutions. This hands-on experience is invaluable and prepares you for tackling similar challenges in your future career.

7 Future work scope

1. Advanced Techniques: Explore deep learning, reinforcement learning, or ensemble methods.
2. Additional Data Sources: Incorporate satellite imagery, remote sensing data, or crop-specific variables.
3. Spatial and Temporal Analysis: Assess regional variations and trends over time.
4. Seasonal Forecasting: Develop models for seasonal crop yield forecasting.
5. Crop-specific Models: Build tailored models for different crops.
6. Economic Factors: Integrate market prices, input costs, and policies into the model.
7. User Interface: Create a user-friendly dashboard for stakeholders.
8. Field Validation: Validate model predictions with real-world data.
9. Scalability: Scale up deployment and coverage.
10. Long-term Impact: Evaluate the model's impact on agricultural productivity and sustainability.