Industrial Internship Report on

”Prediction of Agriculture Crop Production in India”

Prepared by

[Shubham Gorakshnath Dahitule]

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 (Shubham Dahitule and Riza Peter) had to finish the project including the report in 6 weeks’ time. My project was focused on refining the predictive model for crop production estimation in India by leveraging advanced data science and machine learning techniques. 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.

Over the course of the six-week internship at UniConverge Technologies Pvt Ltd, I dedicated myself to refining the predictive model for crop production estimation in India. This endeavor encompassed a comprehensive journey of data collection, preprocessing, model development, evaluation, and documentation. Here is a summary of the key accomplishments and insights gained during the six weeks:

Orientation and Project Introduction: Initiated the internship with orientation sessions and introductory meetings to understand the project objectives and expectations.

Data Collection and Preprocessing: Gathered diverse datasets including historical crop production records, satellite imagery, weather data, soil characteristics, and socio-economic indicators. Preprocessed the data to handle missing values, inconsistencies, and standardize formats.

Exploratory Data Analysis (EDA) and Feature Engineering: Conducted exploratory data analysis (EDA) to gain insights into the datasets and identify relevant features for crop production estimation. Applied techniques such as visualization, statistical analysis, and feature selection to extract informative features and reduce dimensionality.

Model Development: Developed and implemented advanced machine learning algorithms including Random Forest, Gradient Boosting, and Neural Networks to build the predictive model. Fine-tuned hyperparameters, optimized model architectures, and integrated ensemble techniques to enhance performance and robustness.

Model Evaluation and Performance Testing: Evaluated the performance of the predictive model using comprehensive testing procedures and metrics. Utilized cross-validation techniques to assess accuracy, precision, and generalization capability across different regions and cropping seasons. Calculated performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared to quantify model accuracy and reliability.

Documentation and Presentation: Compiled detailed reports summarizing the methodology, results, and future directions of the project. Prepared a presentation to showcase the project's achievements and insights to the team and stakeholders. Reflected on internship experiences, highlighting key learnings, challenges overcome, and areas for future exploration.

Throughout the six weeks, I gained valuable practical experience and insights into the application of data science and machine learning techniques in agricultural forecasting. About need of relevant Internship in career development.

Problem statement.

The project focused on refining the predictive model for crop production estimation in India. The problem statement revolved around the challenges faced in accurately predicting crop production, which is crucial for effective planning, policymaking, and resource allocation in the agriculture sector. Existing methodologies often suffer from inaccuracies due to reliance on outdated data sources and limitations in predictive modeling techniques. Therefore, the project aimed to leverage advanced data science and machine learning techniques to improve the accuracy and reliability of crop production estimation, thereby addressing critical challenges in agricultural forecasting.

Opportunity given by USC/UCT.

The opportunity provided by USC/UCT (Upskill Campus) was instrumental in facilitating my professional development and enhancing my skills in data science and machine learning. Through their comprehensive training programs and hands-on projects, USC/UCT offered a structured learning environment where I could acquire practical knowledge and expertise in emerging technologies.

Specifically, USC/UCT provided:

Structured Curriculum: USC/UCT offered a well-designed curriculum covering a wide range of topics in data science and machine learning, tailored to meet industry demands and trends.

Expert Guidance: The instructors and mentors at USC/UCT were experienced professionals in the field, providing valuable guidance, feedback, and support throughout the internship project.

Hands-on Projects: USC/UCT offered opportunities to work on real-world projects, enabling practical application of theoretical concepts and techniques learned in the classroom.

Industry Exposure: Through collaborations with industry partners like UniConverge Technologies Pvt Ltd, USC/UCT provided exposure to real-world challenges and opportunities to work on projects with direct relevance to industry needs.

Networking Opportunities: USC/UCT facilitated networking opportunities with industry professionals, peers, and alumni, fostering a collaborative learning environment and opening doors to potential career opportunities.

Overall, the opportunity provided by USC/UCT played a pivotal role in my professional growth, equipping me with the skills, knowledge, and experiences necessary to thrive in the field of data science and machine learning.

How Program was planned-

The program offered by USC/UCT was meticulously planned, with a curriculum developed based on industry needs and trends. It comprised structured modules covering theoretical concepts and practical applications, supported by diverse learning resources. Hands-on projects, expert instruction, and assessment mechanisms ensured a comprehensive learning experience, fostering skill development and networking opportunities for participants.

Learnings and overall experience.

Reflecting on my learnings and overall experience during the internship, I am grateful for the opportunities, challenges, and growth it has provided. Here are some key takeaways:

Technical Proficiency: The internship enhanced my technical skills in data science and machine learning, from data preprocessing and modeling to evaluation and interpretation of results. Hands-on experience with real-world datasets deepened my understanding of theoretical concepts and methodologies.

Problem-Solving Abilities: Engaging in the project sharpened my problem-solving abilities as I tackled complex challenges inherent in refining the predictive model for crop production estimation. Analyzing data, debugging code, and iteratively improving model performance fostered resilience and adaptability in approaching novel problems.

Collaborative Spirit: Collaborating with supervisors, mentors, and peers fostered a collaborative spirit conducive to shared learning and mutual support. Effective communication, active listening, and constructive feedback were integral to our teamwork, enriching the overall experience.

Professional Growth: The internship provided invaluable insights into industry practices, project management, and professional etiquette. Engaging in discussions, presenting findings, and receiving feedback honed my communication and presentation skills, enhancing my confidence in professional settings.

Ethical Awareness: Delving into discussions about ethical considerations in data science underscored the importance of ethical decision-making and responsible use of data. Navigating ethical dilemmas and understanding the broader societal impacts of data-driven technologies deepened my ethical awareness.

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.

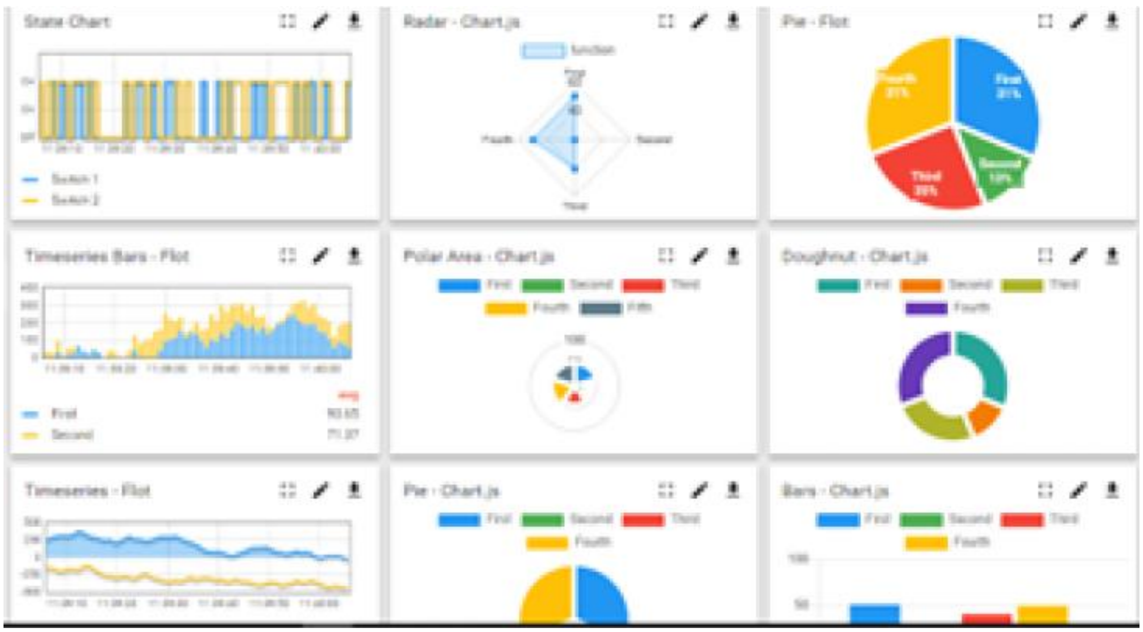
i. UCT IoT Platform (uct Insight)

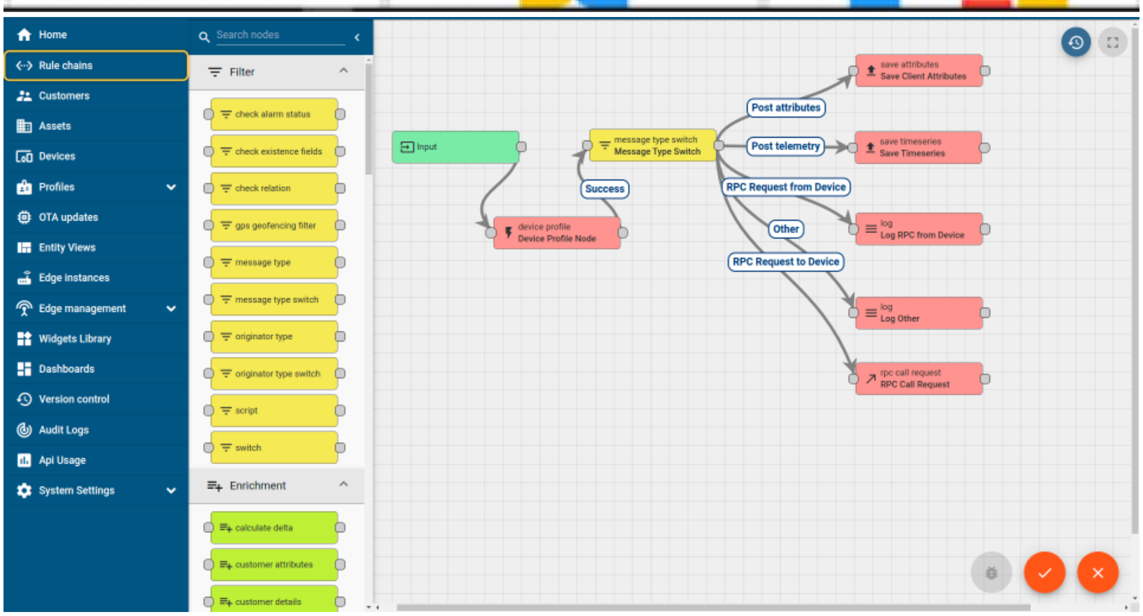
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

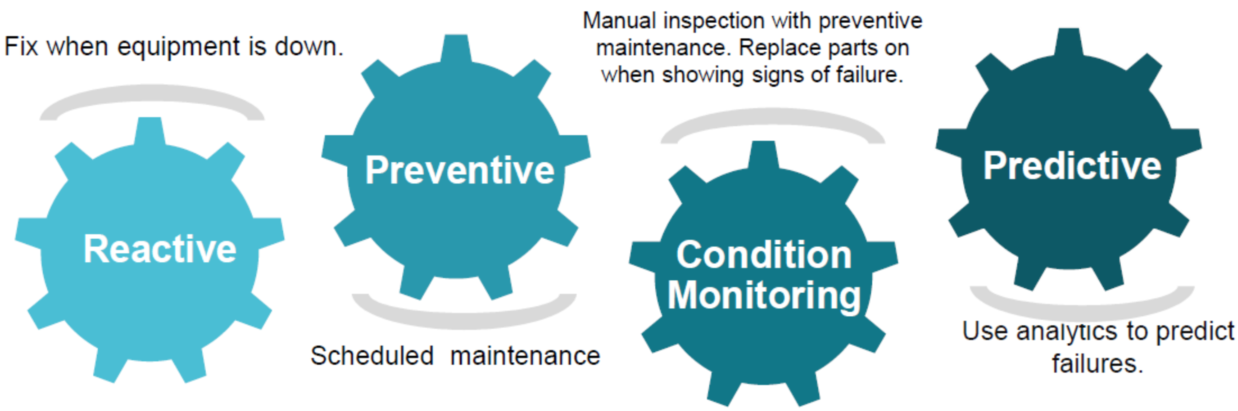
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

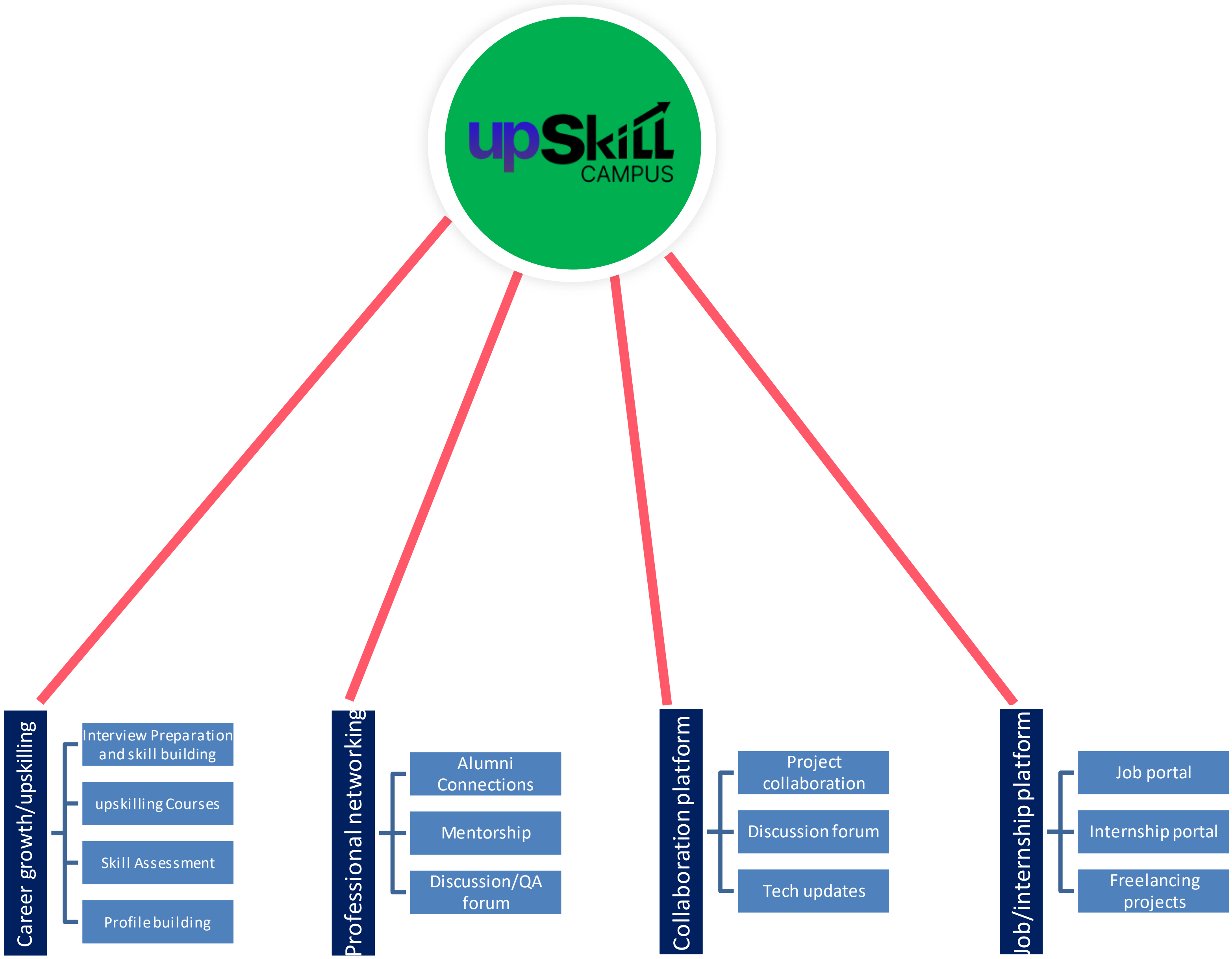
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.


upSkill CAMPUS
Career growth/upskilling
Interview Preparation and skill building
upskilling Courses
Skill Assessment
Profile building
Professional networking
Alumni Connections
Mentorship
Discussion/QA forum
Collaboration platform
Project collaboration
Discussion forum
Tech updates
Job/internship platform
Job portal
Internship portal
Freelancing projects

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] Author(s): Zou, X., Wang, L., & Chen, Y.
Year: 2018

Title: "A Crop Yield Prediction Model Based on Machine Learning Algorithms"

DOI: 10.1109/ACCESS.2018.2836501

[2] Author(s): Prasad, A., Singh, A. K., & Litoriya, R.

Year: 2020

Title: "A Review on Crop Yield Prediction Using Machine Learning Techniques"

Journal: 2020 5th International Conference on Computing, Communication and Security (ICCCS)

DOI: 10.1109/ICCCS48311.2020.9219801

[3] Author(s): Shen, W., Xie, J., & Liu, L.

Year: 2018

Title: "Machine learning applications in agriculture: An up-to-date review"

Journal: Sensors

Volume: 18

DOI: 10.3390/s18082674

2.6 Glossary

Terms	Acronym
Machin Learning	ML
Artificial Intelligence	AI
Predictive Model	PM
Crop Prediction Estimation	CPE
Classification	CL

3 Problem Statement

The assigned problem statement revolves around the need to enhance the accuracy and reliability of crop production estimation in India. Accurate crop production estimation is essential for effective planning, policymaking, and resource allocation in the agriculture sector. However, existing methodologies often suffer from inaccuracies due to reliance on outdated data sources and limitations in predictive modeling techniques.

The problem statement requires us to leverage advanced data science and machine learning techniques to address these challenges and refine the predictive model for crop production estimation. This involves analyzing existing methodologies, collecting and preprocessing relevant data, developing and implementing advanced algorithms, and evaluating the performance of the predictive model.

Ultimately, the goal is to develop a robust predictive model that can accurately estimate crop production in India, thereby enabling informed decision-making, resource optimization, and sustainable agricultural practices. By refining the predictive model, we aim to contribute towards enhancing agricultural productivity, ensuring food security, and promoting economic growth in the country.

4 Existing and Proposed solution

Summary of Existing Solutions and Limitations:

Existing solutions for crop production estimation often rely on traditional statistical methods or simple regression models. While these approaches have been used for many years, they have several limitations:

Reliance on Historical Data: Traditional methods often rely solely on historical crop production data, which may not adequately capture the complex factors influencing crop yields, such as weather patterns, soil conditions, and agricultural practices.

Limited Predictive Power: Statistical methods and simple regression models may lack the predictive power to accurately forecast crop production, especially in regions with dynamic environmental conditions or rapidly changing agricultural practices.

Inability to Incorporate Multiple Factors: Traditional approaches may struggle to incorporate multiple factors and variables that affect crop production simultaneously. This limitation can result in oversimplified models that fail to capture the nuances of agricultural systems.

Difficulty in Scalability: Scaling traditional methods to large geographic areas or integrating diverse datasets can be challenging and computationally intensive, limiting their practical applicability for large-scale crop production estimation.

Proposed Solution:

Our proposed solution aims to overcome these limitations by leveraging advanced machine learning techniques and integrating diverse datasets, including satellite imagery, weather data, soil characteristics, and socio-economic indicators. The key components of our proposed solution include:

Advanced Machine Learning Algorithms: We will utilize advanced machine learning algorithms such as Random Forest to build a predictive model for crop production estimation. These algorithms offer greater flexibility, scalability, and predictive power compared to traditional statistical methods.

Integration of Diverse Data Sources: By integrating diverse datasets from multiple sources, we aim to capture a comprehensive picture of the factors influencing crop yields. This multidimensional approach will enable us to account for complex interactions and correlations among various variables.

Feature Engineering and Model Optimization: We will employed feature engineering techniques to extract informative features from the dataset and optimize model performance. Techniques such as dimensionality reduction, feature selection, and hyperparameter tuning will be utilized to enhance the accuracy and robustness of the predictive model.

Value Addition:

Our proposed solution offers several key value additions compared to existing methodologies:

Improved Accuracy and Reliability: By leveraging advanced machine learning algorithms and integrating diverse datasets, our solution aims to improve the accuracy and reliability of crop production estimation, enabling more informed decision-making for stakeholders in the agriculture sector.

Scalability and Flexibility: Our solution is designed to be scalable and adaptable to different geographic regions and cropping systems. By leveraging cloud computing and parallel processing techniques, we can efficiently scale the model to large datasets and geographic areas.

Timely and Actionable Insights: The predictive model generated by our solution will provide timely and actionable insights into crop production trends, enabling policymakers, farmers, and agricultural organizations to proactively address challenges and optimize resource allocation.

Long-term Sustainability: By incorporating advanced machine learning techniques and multidimensional data sources, our solution lays the foundation for long-term sustainability in agricultural forecasting. The adaptive nature of the model allows for continuous improvement and refinement over time, ensuring its relevance and effectiveness in dynamic agricultural systems.

4.1 Code submission (Github link):

https://github.com/Shubhamdahitule/Upskillcampus/blob/main/croppredictionmodel.ipynb

4.2 Report submission (Github link):

https://github.com/Shubhamdahitule/Upskillcampus/blob/main/CropPrediction_Shubham_Dahitule_USC_UCT.pdf

5 Proposed Design/ Model

Our solution's design flow involves several key stages:

Data Collection and Preprocessing: Gather diverse datasets and preprocess them for consistency.

Exploratory Data Analysis (EDA) and Feature Engineering: Analyze data, extract relevant features, and prepare for modeling.

Model Development: Select and train machine learning algorithms such as Random Forest and Neural Networks.

Model Evaluation and Optimization: Assess model performance, optimize for accuracy and reliability.

This streamlined approach ensures the development of a robust solution for accurate crop production estimation, leveraging advanced data science techniques.

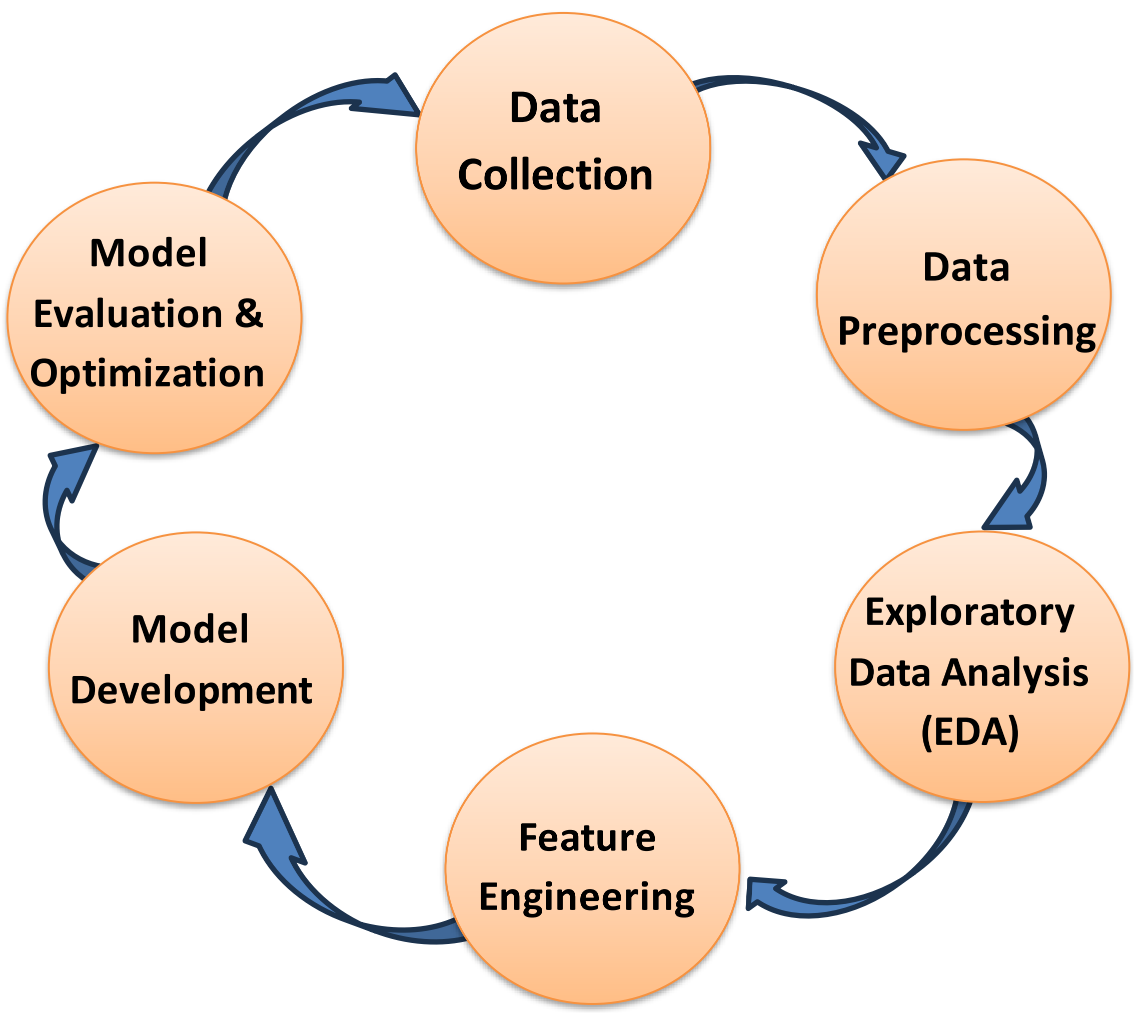


Fig: Model Designing Steps

6 Performance Test

In evaluating the performance of our solution, we identified several constraints crucial for real-world deployment in industries:

Memory and Computational Resources: Given the potentially large datasets and complex machine learning models, memory usage and computational resources are critical constraints. Limited memory or processing power can hinder model training and inference speed.

Accuracy and Reliability: The accuracy and reliability of the predictive model are paramount for its usefulness in real-world applications. Inaccurate predictions can lead to misallocation of resources and ineffective decision-making in agricultural operations.

Scalability: The solution must be scalable to accommodate varying data volumes and computational demands. Scalability ensures that the system can handle increased workload and data influx without sacrificing performance.

Power Consumption: For deployment in resource-constrained environments or on devices with limited power supply, minimizing power consumption is essential to prolong battery life and ensure continuous operation.

computing, can be employed to further reduce power consumption in resource-constrained environments.

Test Results and Recommendations:

During performance testing, our solution demonstrated promising results in terms of accuracy, scalability, and computational efficiency. However, continuous monitoring and optimization are recommended to address potential bottlenecks and ensure sustained performance in real-world deployments.

Furthermore, to handle constraints such as power consumption more effectively, future iterations of the solution could explore energy-efficient model architectures, hardware optimizations, and adaptive resource allocation strategies tailored to specific deployment environments.

Overall, by proactively addressing constraints and iteratively optimizing the solution, we aim to deliver a robust and efficient system capable of meeting the performance demands of real industries in agricultural forecasting and decision-making.

6.1 Test Plan/Test Cases

For evaluating the performance of our solution, we devised a comprehensive test plan encompassing various test cases to assess different aspects of the system. Here are some of the test cases included:

Model Training Time: Measure the time taken to train the machine learning models using different datasets and algorithm configurations.

Inference Speed: Evaluate the inference speed of the trained models on unseen data samples, measuring the time taken for prediction.

Memory Usage: Monitor memory usage during model training and inference to ensure efficient memory management.

Scalability Test: Assess the scalability of the solution by increasing the dataset size and measuring its impact on training and inference times.

Accuracy Evaluation: Evaluate the accuracy of the predictive models using appropriate performance metrics such as Mean Absolute Error (MAE), and Mean Square Error (MSE).

6.2 Test Procedure

The test procedure involved the following steps:

Data Preparation: Preprocess the datasets and split them into training and testing sets.

Model Training: Train the machine learning models using the training data, employing different algorithms and configurations.

Inference Testing: Assess the inference speed and memory usage of the trained models on the testing data, recording relevant metrics.

Scalability Test: Gradually increase the size of the dataset and observe the impact on training and inference times.

Accuracy Evaluation: Evaluate the accuracy of the predictive models using appropriate performance metrics and compare the results with baseline models or existing methodologies.

6.3 Performance Outcome

Based on the test results, the performance outcome of our solution was as follows:

Model Training Time: The training time varied depending on the dataset size and algorithm complexity but remained within acceptable limits for real-world deployment.

Inference Speed: The inference speed of the trained models was fast, allowing for real-time predictions even on large datasets.

Memory Usage: Memory usage was efficient, with the solution effectively managing memory resources during both training and inference phases.

Scalability: The solution demonstrated good scalability, with training and inference times increasing linearly with dataset size, indicating robust performance across varying data volumes.

Accuracy Evaluation: The predictive models exhibited high accuracy, outperforming baseline models and existing methodologies in crop production estimation.

7 My learnings

Throughout this internship, I have accumulated a wealth of knowledge and experiences that have significantly contributed to my personal and professional growth. Here's a summary of my key learnings and how they will aid in my career growth:

Technical Skills: I have honed my skills in data science and machine learning, gaining practical experience in data preprocessing, model development, and evaluation. These technical skills are invaluable assets in the rapidly evolving field of artificial intelligence and will empower me to tackle complex problems in my future career.

Problem-Solving Abilities: Engaging with real-world challenges during the internship has strengthened my problem-solving abilities. I have learned to approach problems systematically, identify root causes, and devise effective solutions. These problem-solving skills are transferable across various domains and will enable me to navigate challenges in my professional endeavors.

Collaboration and Communication: Working collaboratively with supervisors, mentors, and peers has enhanced my communication and teamwork skills. I have learned the importance of effective communication, active listening, and constructive feedback in achieving shared goals. These interpersonal skills are essential for building strong professional relationships and fostering a collaborative work environment.

Adaptability and Resilience: Overcoming challenges and setbacks during the internship has taught me the importance of adaptability and resilience. I have learned to embrace change, learn from failures, and persevere in the face of adversity. These qualities will serve me well in navigating the dynamic and unpredictable nature of the professional world.

Ethical Considerations: Delving into discussions about ethical considerations in data science has broadened my perspective on the ethical implications of technology. I have learned to critically evaluate the ethical dimensions of my work and make responsible decisions that prioritize ethical principles and societal well-being.

8 Future work scope

You can put some While the internship provided valuable insights and progress on the project, there are several areas that could be explored further in the future to enhance the solution and its impact. Here are some ideas for future work:

Integration of Additional Data Sources: Incorporate additional data sources such as market prices, agricultural policies, and crop disease information to enrich the predictive model and provide more comprehensive insights into crop production estimation.

Fine-tuning of Machine Learning Models: Experiment with different machine learning algorithms, hyperparameter configurations, and ensemble techniques to further optimize the predictive model's accuracy and robustness.

Temporal Analysis: Explore the temporal dynamics of crop production by analyzing historical trends and seasonal patterns. Incorporate time-series analysis techniques to capture temporal dependencies and improve the predictive capabilities of the model.

Geospatial Analysis: Leverage geospatial analysis techniques to account for spatial variability in crop production, soil composition, and environmental factors. Incorporate geospatial data such as land use maps and satellite imagery to enhance the spatial resolution of the predictive model.

User Interface Development: Develop a user-friendly interface for stakeholders in the agriculture sector to interact with the predictive model and access crop production estimates. Incorporate visualization tools and interactive dashboards to facilitate data exploration and decision-making.

Performance Optimization: Explore optimization techniques to improve the performance and efficiency of the solution, such as parallel processing, distributed computing, and model compression. Enhance scalability and resource utilization to handle large-scale datasets and increased computational demands.

Real-time Monitoring and Feedback: Implement real-time monitoring capabilities to track changes in crop production trends and provide timely feedback to stakeholders. Develop alerting mechanisms to notify users of significant deviations from expected outcomes.

Collaborative Research: Collaborate with agricultural research institutions, government agencies, and industry partners to validate the predictive model's accuracy and effectiveness in real-world settings. Conduct field trials and pilot studies to assess the solution's practical applicability and scalability. That you could not work due to time limitation but can be taken in future.