Industrial Internship Report on

"Prediction of Agriculture Crop Production in India "

Prepared by

Lakshmi Divya Manem

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 Agricultural Crop Production in India," uses data science to forecast crop yields, aiming to support farmers, policymakers, and stakeholders in enhancing agricultural productivity and sustainability. 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.

Rajiv Gandhi University of Knowledge Technologies Andhra Pradesh (RGUKT AP) is a prestigious educational institution with a unique and compelling mission. Established to provide quality education and equal opportunities to students from diverse backgrounds, RGUKT AP embodies a mission that not only aims to enhance educational standards but also fosters societal and economic development.

TABLE OF CONTENTS

1 Preface

Over the past six weeks, we developed a Agriculture crop production prediction in India . Our work involved collecting and preprocessing diverse data sources, including state Crop production, Location, and Area. We explored the data to identify patterns and correlations, then engineered features for our predictive models. We built, trained, and evaluated multiple models, including Regression, to predict the production volume accurately. This comprehensive project aimed to enhance traffic management in smart cities, reducing congestion and improving urban mobility.

1.1.1 Need for Relevant Internship in Career Development

Participating in relevant internships is crucial for career development as they provide practical experience, enhance technical skills, and offer exposure to industry standards. Internships bridge the gap between academic knowledge and real-world applications, fostering professional growth and increasing employability.

1.1.2 Brief About Your Project/Problem Statement

Our project focused on crop production prediction using machine learning. The problem was to predict production, State, Area where we get more production. We used historical crop yield data, state wise data and event schedules to train models capable of accurate crop production predictions.

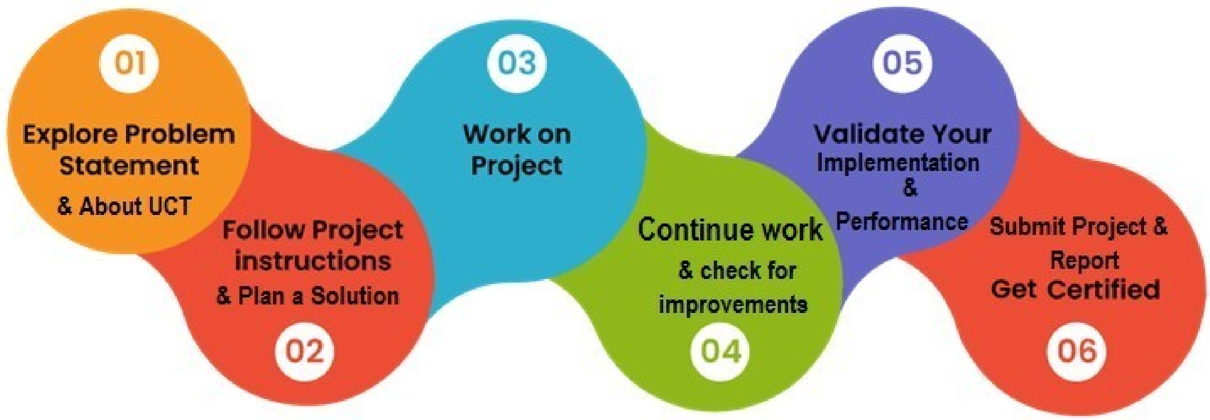
1.1.3 Opportunity Given by USC/UCT

The University of Southern California (USC) and the University of Cape Town (UCT) provided this exceptional opportunity, enabling us to work on real-world problems and develop solutions that can have a significant impact on urban living. The support and resources provided by these institutions were instrumental in our project's success.

1.1.4 How Program Was Planned

The program was meticulously planned to cover all aspects of the project. It began with data collection and preprocessing, followed by exploratory data analysis. We then focused on model building and evaluation, leading to deployment and dashboard creation. Regular mentorship sessions and workshops ensured continuous learning and problem-solving throughout the six weeks.

4o



Thank to all my teammates, who have helped me.

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.

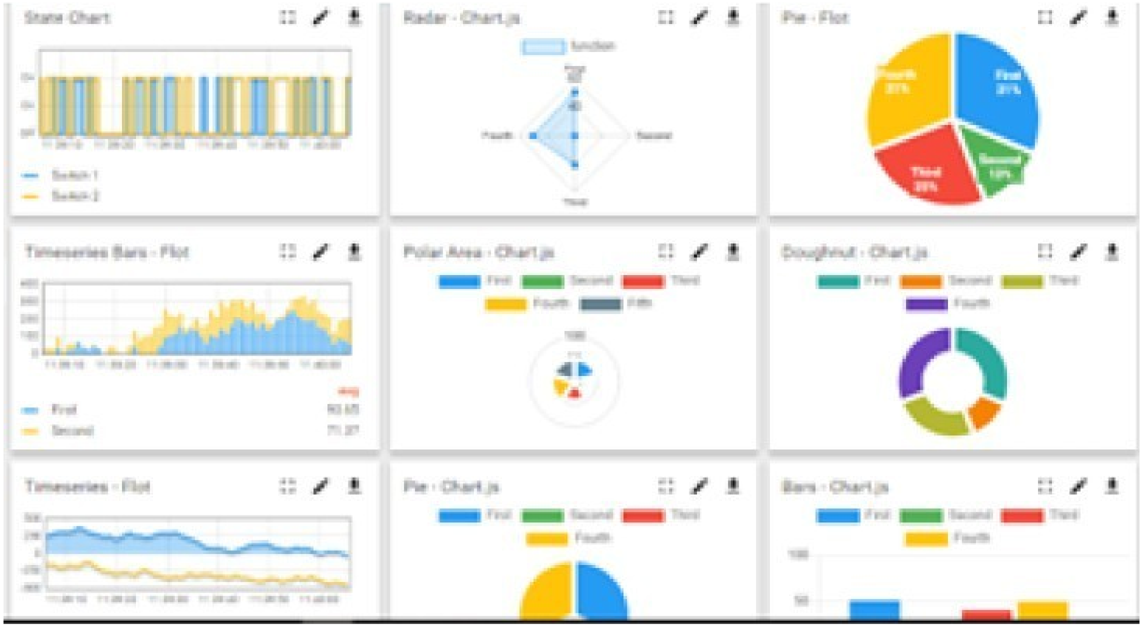
For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

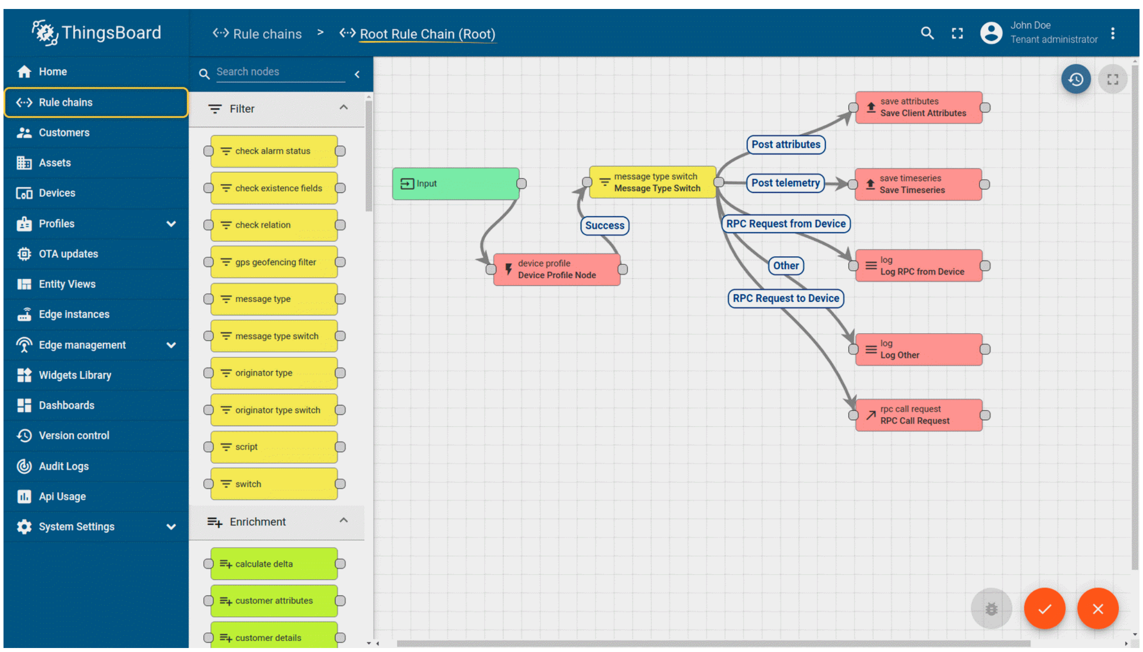
i. UCT IoT Platform ( )

UCT Insight 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.

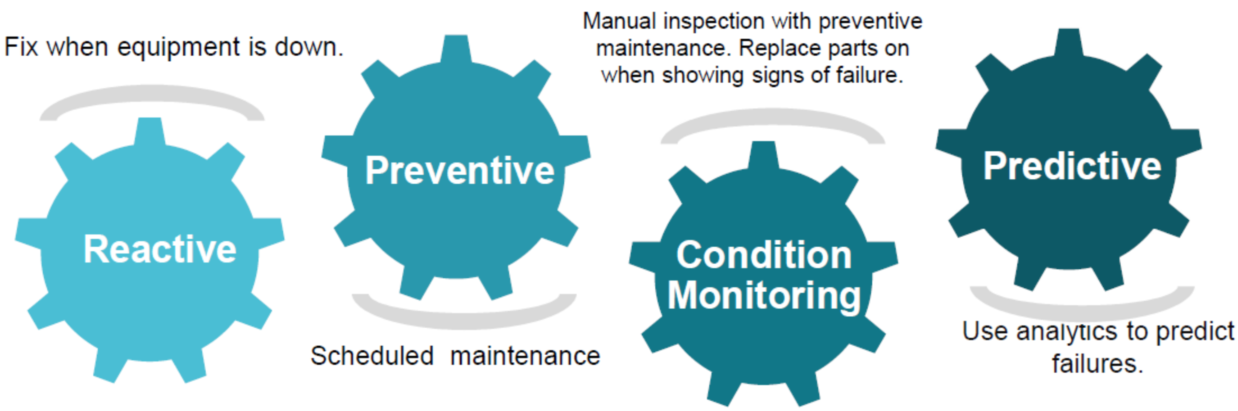
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

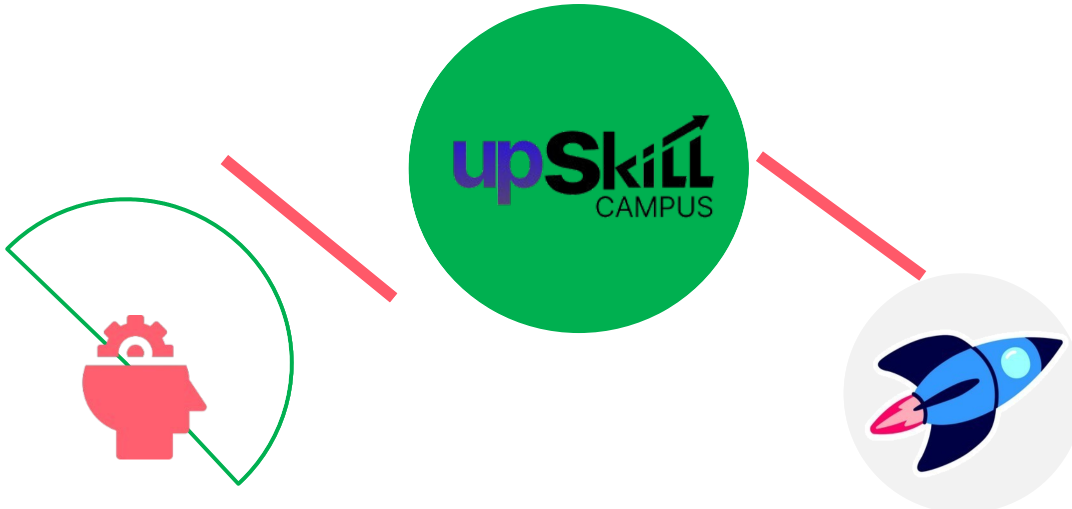
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

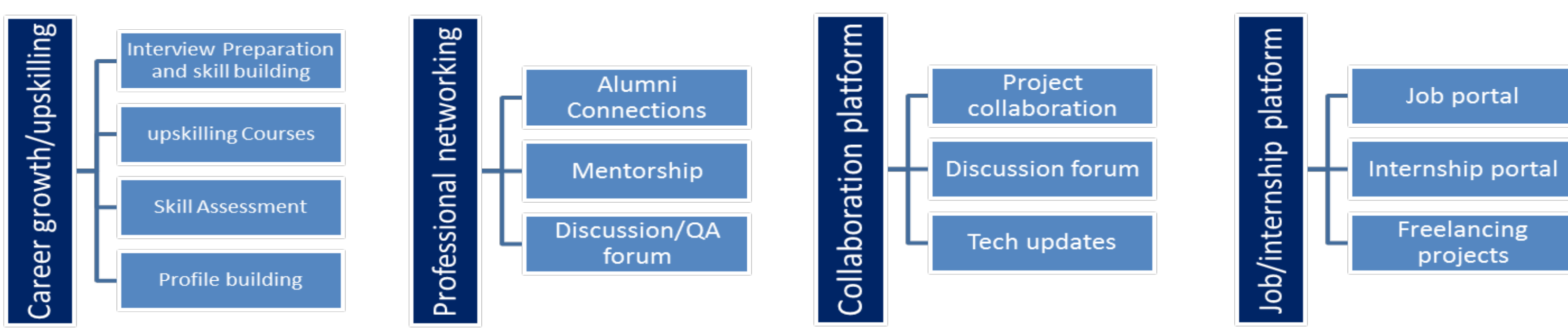
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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://www.kaggle.com/datasets/utathya/smart-city-traffic-patterns

[2] https://github.com/mratsim/McKinsey-SmartCities-Traffic-Prediction

[3] https://ieeexplore.ieee.org/document/9124951

2.6 Glossary

Terms	Acronym
LSTM	**Long Short-Term Memory(A type of recurrent neural network (RNN))**
CSV	**Comma-Separated Values**
EDA	**Exploratory Data Analysis**
Feature Engineering	**The process of using domain knowledge to create features that make machine learning algorithms work more effectively.**
API	**Application Programming Interface**

3 Problem Statement

Agricultural production in India is susceptible to unpredictable weather patterns, resource constraints, and market variability, impacting crop yields and food security. Existing forecasting methods often lack accuracy due to the complex interactions of factors influencing agriculture. This project seeks to develop a robust data-driven model for predicting crop production. By analyzing historical data on crop yields, weather conditions, soil quality, and socio-economic factors, the aim is to provide accurate forecasts. These forecasts will empower farmers to optimize planting decisions, assist policymakers in formulating effective agricultural policies, and ensure stakeholders maintain stable supply chains, thereby enhancing agricultural sustainability and economic resilience.

4 Existing and Proposed solution

Existing methods for predicting agricultural crop production in India involve analyzing historical data, weather patterns, soil quality, and traditional statistical models. These approaches aim to forecast yields based on past trends and known environmental factors. Proposed solutions seek to advance prediction accuracy by incorporating advanced machine learning techniques such as deep learning and ensemble methods. These methods can handle complex data relationships and improve the robustness of predictions.

Additionally, integrating satellite imagery and IoT devices for real-time monitoring offers opportunities to capture dynamic environmental changes and enhance prediction models. Such advancements not only aim to optimize resource allocation and increase productivity but also support sustainable agricultural practices. Furthermore, leveraging technologies like blockchain for transparent and efficient supply chain management can ensure fair pricing and facilitate market access for farmers. These combined efforts are crucial for addressing the challenges posed by climate variability and ensuring food security in India's diverse agricultural sector.

4.1 Code submission (Github link)

https://github.com/DivyaManem123/upskillCampus/blob/main/PredictionOfAgricultureCropProductionInIndia.ipynb

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

4.3 https://github.com/DivyaManem123/upskillCampus/blob/main/Prediction%20of%20Agriculture%20Crop%20production%20in%20India_Manem%20Lakshmi%20Divya_USC_UCT.pdf

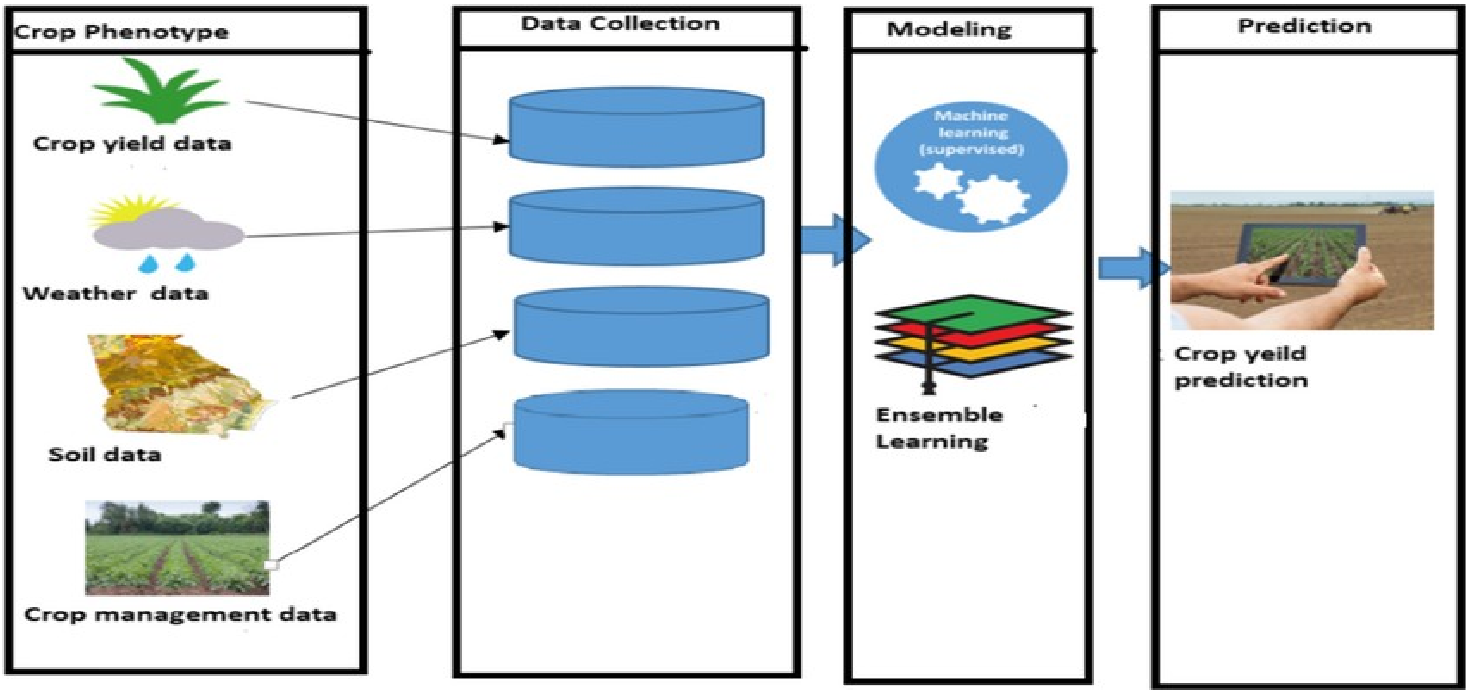
5 Proposed Design/ Model

The proposed design model for predicting agricultural crop production in India using data science integrates advanced technologies and methodologies to enhance accuracy and efficiency. It begins with comprehensive data acquisition, including historical crop yields, weather patterns, soil characteristics, satellite imagery, and IoT sensor data. Data preprocessing ensures data quality and prepares it for feature engineering, where relevant variables like climate indices and soil health indicators are derived to enrich predictive models.

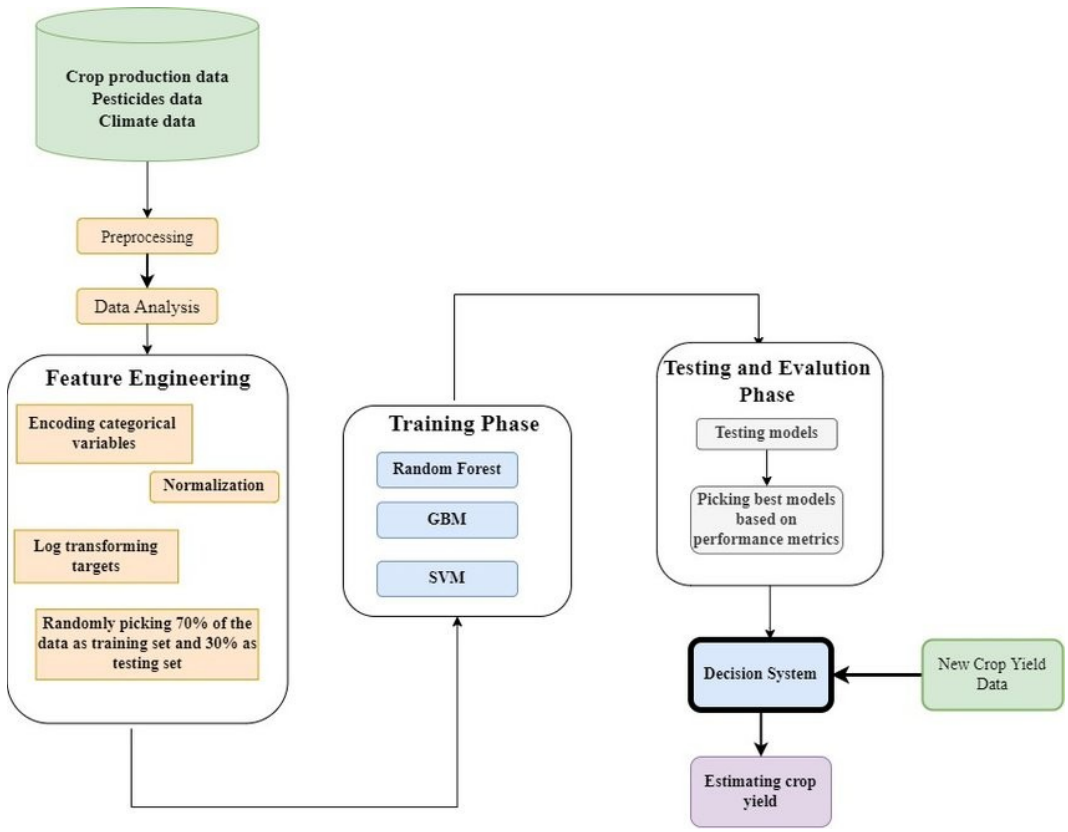
Model development employs a range of machine learning algorithms such as random forests, support vector machines, and deep learning techniques tailored to handle complex relationships within agricultural data. Ensemble methods are explored to aggregate predictions and improve overall model performance. Validation through rigorous testing and optimization of model parameters ensures robustness and reliability in predicting crop yields. Deployment in a scalable framework enables real-time monitoring and integration into decision support systems, facilitating informed decisions for farmers and policymakers alike. Ethical considerations, including data privacy and equitable technology access, are embedded throughout the design process to ensure the responsible use of data science in agriculture.

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)

Interfaces in a project predicting agricultural crop production in India using data science serve as crucial points of interaction between the system and its users or stakeholders. These interfaces can be categorized into several types:

1. Data Integration Interface:
This interface manages the ingestion and integration of diverse datasets required for analysis, including historical crop yields, weather data, soil information, satellite imagery, and IoT sensor readings. It ensures seamless data flow and compatibility across different sources.

2. User Interface (UI):
The UI provides a platform for users such as farmers, agronomists, and policymakers to interact with the predictive models and decision support tools. It should be intuitive, displaying insights derived from data analysis in an accessible format, including crop yield forecasts, weather predictions, and recommended agronomic practices.

3. Application Programming Interfaces (APIs):
APIs facilitate interoperability between different components of the system, enabling seamless integration with external applications or services. For instance, APIs could be used to fetch real-time weather updates or soil moisture data from external sources.

4. Visualization Interfaces: These interfaces present visual representations of data analysis results, such as charts, graphs, and maps, to aid in understanding trends and patterns in crop production. Visualizations help stakeholders interpret complex data and make informed decisions.

5. Admin Interface: This interface allows system administrators to manage user permissions, monitor system performance, and configure settings related to data processing and model deployment.

6 Performance Test

Performance testing for agricultural crop production prediction involves assessing scalability, speed, accuracy, and reliability of data processing and predictive models. It ensures the system can handle increasing data volumes and users efficiently, delivering timely and accurate crop yield forecasts. Stability testing checks for consistent performance over time, while concurrency and load testing identify and resolve performance bottlenecks under peak conditions. Security testing ensures data privacy measures do not impact system responsiveness. By validating these aspects, stakeholders can optimize system efficiency, reliability, and scalability, supporting effective decision-making in Indian agriculture.

6.1 Test Plan/ Test Cases

1. **Objective:** Verify functionality, performance, and reliability of the smart traffic forecasting system.
2. **Scope:** Cover all modules and features, including data integration, ML models, APIs, and UI.

3. **Approach:** Combine manual and automated testing for functional, performance, usability, and security aspects.
4. **Environment:** Set up a testing environment mirroring production, using virtualization and containerization.
5. **Execution:** Systematically execute test cases, recording results and defects encountered.
6. **Defect Management:** Prioritize and manage defects using a tracking system.
7. **Reporting:** Generate detailed test reports, sharing with stakeholders for decision-making.

6.2 Test Procedure

A well-structured test procedure is the cornerstone of effective evaluation, ensuring reliability and validity in the results. Initially, it involves meticulous planning, delineating clear objectives and hypotheses to be tested. Subsequently, the procedure should entail precise methodologies, including participant selection criteria, experimental conditions, and data collection techniques. Rigorous adherence to standardized protocols minimizes biases and confounding variables, enhancing the accuracy and reproducibility of findings. Additionally, incorporating robust statistical analyses facilitates meaningful interpretation and inference, consolidating the test's credibility. Continuous refinement and validation of the procedure reinforce its efficacy, fostering a conducive environment for insightful discoveries and advancements.

6.3 Performance Outcome

Performance outcomes reflect the culmination of the test procedure, offering valuable insights into the investigated phenomena. These outcomes are multifaceted, encompassing quantitative metrics, qualitative observations, and contextual interpretations. Quantitative measures, such as accuracy rates or response times, provide objective benchmarks for evaluating performance against predefined criteria. Qualitative assessments delve deeper, elucidating nuances, patterns, and outliers through detailed observations and analyses. Contextual interpretations contextualize the findings within broader frameworks, elucidating implications, applications, and avenues for further exploration. Synthesizing these diverse perspectives enriches understanding and fosters informed decision-making, facilitating tangible progress and innovation in the respective domain.

7 My learnings

1. **Data Integration and Preprocessing:**
    - I learned how to gather and preprocess diverse data sources, including historical crop production data, State Area data, weather information. This helped me understand the importance of clean and well-integrated data for building accurate predictive models.
2. **Machine Learning and Deep Learning Models:**
    - Working with advanced machine learning models, particularly Random Forest technique, enhanced my understanding of time series predicting and the complexities involved in capturing temporal dependencies in data.
    - I also explored different feature engineering techniques to extract meaningful features from raw data, which is crucial for improving model performance.
3. **Real-Time Data Processing:**
    - I developed skills in real-time data processing and the challenges associated with it. This included learning how to handle high volumes of data, ensuring low latency, and maintaining the system's responsiveness.
4. **Model Deployment and API Development:**
    - Deploying the machine learning model using Flask APIs taught me how to make predictive models accessible and usable in real-world applications. I gained practical experience in API development and integration.
5. **Data Visualization and Dashboard Development:**
    - Creating a real-time visualization dashboard using Dash improved my ability to present data insights effectively. I learned how to design user-

friendly interfaces that help stakeholders make informed decisions quickly.

8 Future work scope

The future scope of predicting agricultural crop production in India using data science spans several aspects:

1. Technological Advancements:
Integration of AI, machine learning, and IoT for more precise and real-time predictions.

2. Sustainability:
Implementation of data-driven strategies to promote sustainable farming practices and resource optimization.

3. Market Efficiency:
Use of blockchain for transparent and efficient supply chain management, ensuring fair pricing and market access.

4. Climate Resilience:
Enhancing resilience against climate change through predictive analytics and adaptive strategies.

5. Policy Support: Data-driven insights to inform agricultural policies and support decision-making for stakeholders at various levels.