Industrial Internship Report on

"Prediction of Agriculture Crop Production in India"

Prepared by

Sheefa Naaz

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 aimed to forecast crop yields in India using machine learning regression models. Through a comprehensive dataset encompassing historical crop production, weather, and soil data, four regression models (Decision Tree, Linear Regression, XGBoost, and Random Forest) were evaluated with MSE and MAE metrics. The XGBoost model exhibited superior performance, achieving an MSE of 113.12 and an MAE of 6.52, showcasing its accuracy in crop yield prediction. This model presents valuable insights for stakeholders to optimize resource allocation and foster sustainable agricultural practices. The project's findings hold significant potential for India's agricultural sector, offering data-driven guidance for farmers, policymakers, and stakeholders to enhance productivity and decision-making. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

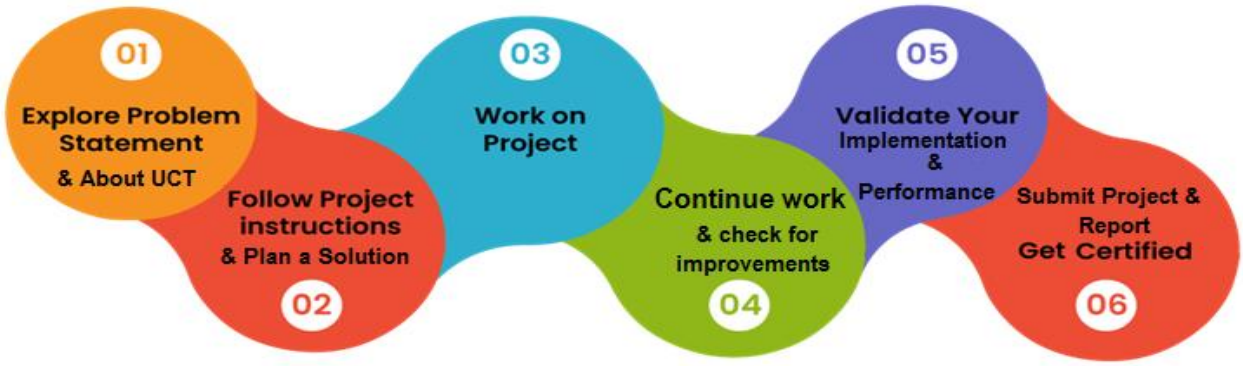
During my six-week internship at UniConverge Technologies Pvt. Ltd, I explored key components such as IoT Academy, Upskill Campus, and UniConverge Technologies Pvt. Ltd. I grasped the significance of machine learning and data science in various industries. The project on "Prediction of Agriculture Crop Production in India" deepened my understanding of Big Data's impact and the roles of data scientists and data analysts. Equipped with essential tools and knowledge, I refined relevant data, implemented statistical analysis, and machine learning models to forecast crop yields. I focused on honing my data scientist skill-set, preparing for interviews, and concluded with a comprehensive project report encompassing data analysis and crop production predictions.

Relevant internships are crucial for career development, offering hands-on learning, industry insights, skill development, networking opportunities, resume enhancement, and personal growth. They align career goals, build strong references, and cultivate essential life skills, contributing to a successful and fulfilling career journey.

Agriculture plays a pivotal role in India's economy, with millions of livelihoods depending on successful crop yields. However, ensuring consistent and optimal crop production remains a significant challenge due to various factors, including climate variability, resource constraints, and changing agricultural practices. The primary goal of this project is to create predictive models that can estimate crop yields in different regions of India based on historical production data and relevant agricultural parameters. We will explore multiple regression algorithms, including Decision Trees, Linear Regression, XGBoost, and Random Forest, to find the most suitable approach for crop yield prediction.

USC/UCT provided me valuable opportunities for career growth and skill development in data science and machine learning. Their internship programs offer hands-on experience on real-world projects, while mentorship and guidance nurture talent. Employees benefit from exposure to diverse industries and innovative projects, contributing to personal and professional growth. USC/UCT's focus on skill development, networking, and career advancement makes it an ideal platform for aspiring professionals to excel in their fields and make meaningful contributions to the tech industry.

The program was meticulously planned, aligning goals with the organization's vision. Audience needs were assessed, and a tailored curriculum was developed. Expert trainers, interactive elements, and assessments were included for an impactful learning experience. Flexibility and feedback mechanisms ensured continuous improvement.



My internship at UniConverge Technologies Pvt. Ltd. was a transformative experience. I gained practical knowledge in data science and machine learning while working on real-world projects. The collaborative environment, skill development opportunities, and focus on problem-solving improved my abilities and confidence. I am grateful for the networking opportunities and look forward to applying my learnings in my future career, well-prepared for success in the data science industry.

I would like to express my gratitude to the entire Upskill Campus, The IoT Academy and UniConverge Technologies Pvt. Ltd. team for creating a conducive and motivating work environment. Their constructive feedback and teamwork have been instrumental in overcoming challenges and achieving success.

To my juniors and peers,

As I reflect on my internship experience at UniConverge Technologies Pvt. Ltd., I want to share a few words of encouragement and advice. Embrace every opportunity that comes your way, as each experience, whether big or small, contributes to your growth. Don't be afraid to step out of your comfort zone and take on new challenges. Embrace the learning process wholeheartedly, as it is through mistakes and setbacks that we truly learn and improve.

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)

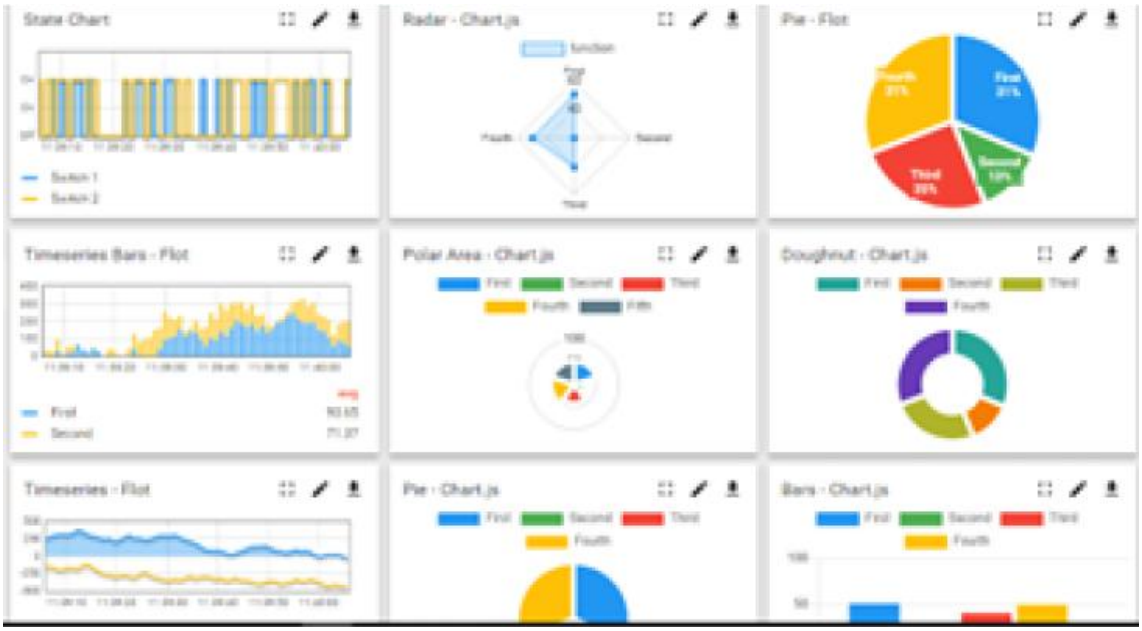
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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

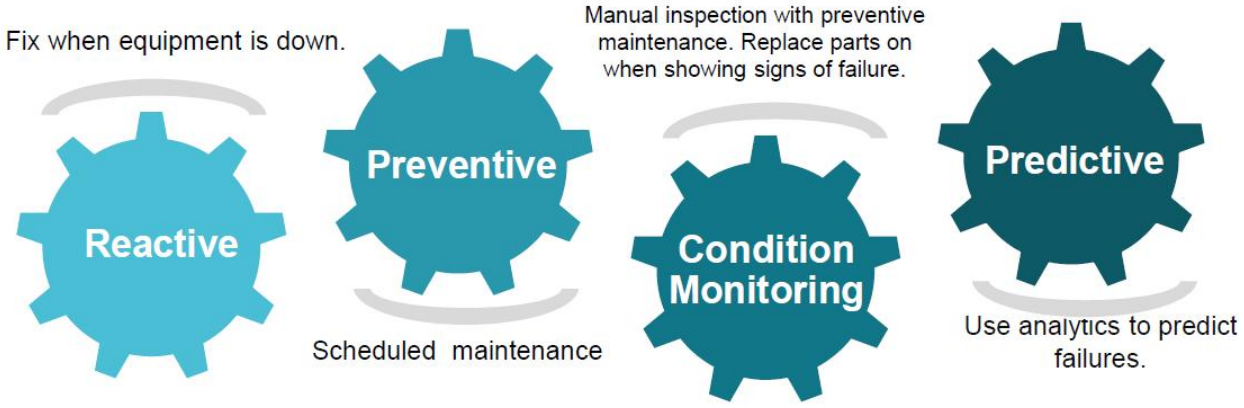
- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular 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.

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

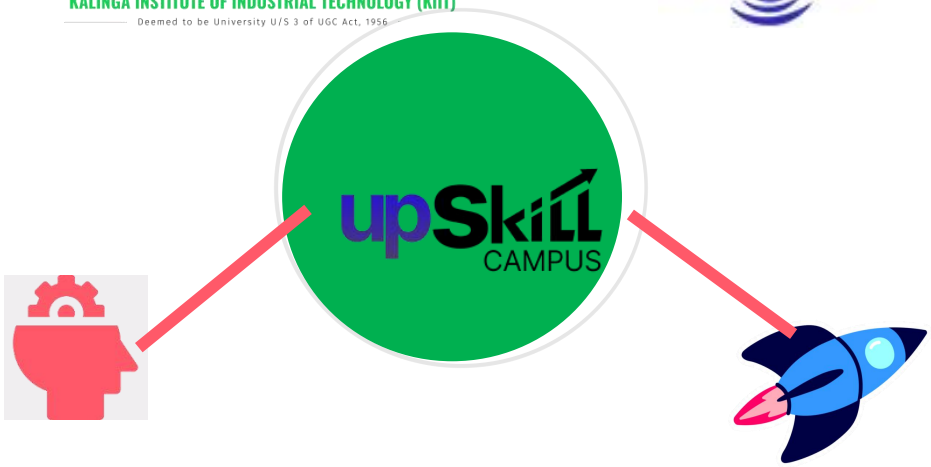
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

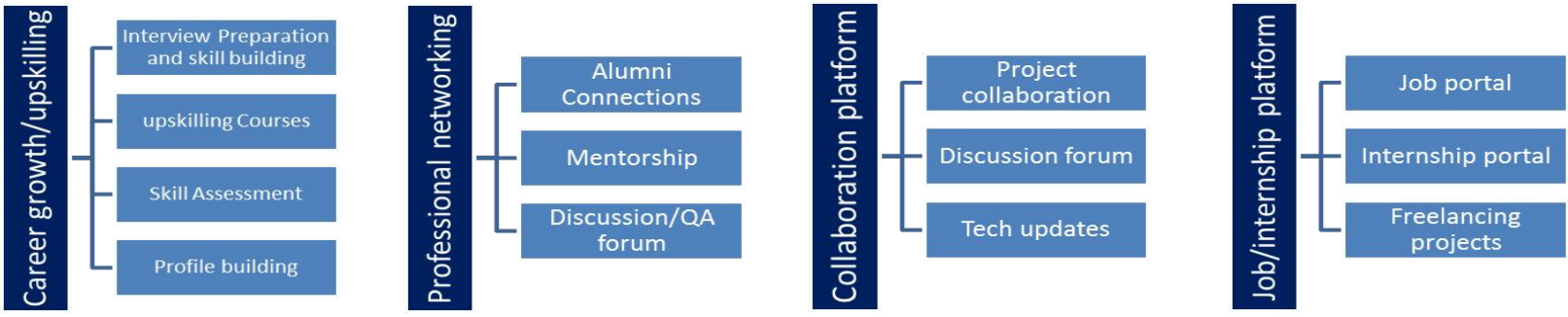
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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.co



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] Davy Cielen, Arno D.B. Meysman, Mohamed Ali. *Introducing Data Science.*

[2] Wiley Series. An Introduction to Probability and Statistics.

[3] Alex Smola, S.V.N Vishwanathan. Introduction to Machine Learning.

[4] The IoT Academy. What are optimization techniques in machine learning?

2.6 Glossary

Terms	Acronym
MSE	Mean Squared Error
MAE	Mean Absolute Error
USC	Upskill Campus
UCT	UniConverge Technologies Pvt. Ltd
IoT	Internet of Things

3 Problem Statement

The agricultural sector is the backbone of India's economy, employing millions of farmers and contributing significantly to the nation's GDP. However, the agricultural industry faces numerous challenges, including unpredictable weather patterns, limited resources, and the need for sustainable practices. To address these challenges, there is a pressing need for accurate predictions of crop production.

The problem at hand is to develop a data-driven predictive model that can forecast agriculture crop production in India with high accuracy. The project will utilize historical agricultural data, encompassing crop yields, weather patterns, soil properties, and other relevant factors, to build machine learning models capable of predicting crop production for different regions and time periods.

The significance of this project lies in its potential to empower farmers, policymakers, and agricultural stakeholders with valuable insights. Accurate predictions of crop production can lead to informed decision-making, optimized resource allocation, and risk management strategies. By anticipating fluctuations in crop yields, farmers can better plan their agricultural activities, leading to improved productivity and profitability.

Moreover, policymakers can use the predictive model to design targeted interventions, such as irrigation schemes and crop insurance policies, to support farmers in times of potential crop failure due to adverse weather conditions. Sustainable farming practices can also be promoted through the identification of optimal cropping patterns based on predicted production.

The project will involve comprehensive data exploration, feature engineering, and model selection to address the challenges associated with crop production prediction. Factors like changing climate patterns, soil health, and other environmental variables will be taken into account to build a robust and accurate predictive model.

Ultimately, the successful development and implementation of this predictive model can play a vital role in shaping the future of Indian agriculture. It has the potential to foster agricultural growth, enhance food security, and uplift the livelihoods of millions of farmers across the nation. As India strives for sustainable and resilient agriculture, the insights derived from this project will contribute significantly to the growth and well-being of the farming community and the nation as a whole.

4 Existing and Proposed solution

Existing solutions for predicting agriculture crop production in India encompass various approaches, including statistical time series models, climate-based models, remote sensing, and machine learning algorithms. Time series models like ARIMA and STL analyze historical crop production data, but may struggle to capture complex relationships and external factors. Climate-based models focus on weather data's influence on crops but may overlook crucial factors like soil health. Remote sensing and satellite imagery monitor vegetation and land usage, but their accuracy can be affected by cloud cover and data availability. Machine learning algorithms offer predictive capabilities but require extensive data preprocessing and tuning. However, these existing solutions possess limitations. Limited data sources and data quality issues may affect their reliability. Overfitting may hinder generalization to unseen data. Seasonal and regional variations pose challenges in predicting diverse agro-climatic conditions. Some complex machine learning models lack interpretability, impacting their usability for decision-making.

To address these limitations, the proposed project aims to develop a comprehensive and accurate predictive model by incorporating diverse data sources, employing robust machine learning algorithms, and emphasizing feature engineering techniques. The objective is to enhance prediction accuracy, supporting informed decision-making for farmers, policymakers, and stakeholders in the agriculture sector.

The project aims to add value by developing an accurate predictive model for crop production in India. Utilizing advanced machine learning algorithms and comprehensive data, it aims to improve precision and empower stakeholders with reliable insights for informed decision-making and resource allocation. The model's interpretability will enhance user-friendliness, ensuring transparency in predictions. The project's value lies in contributing to sustainable agricultural practices, food security, and overall sector growth in India.

4.1 Code submission (Github link)

https://github.com/sheefanaaz123/UpSkill-Campus/blob/main/PredictionOfAgricultureCropProductionInIndia_SheefaNaaz_USC_UCT.ipynb

4.2 Report submission (Github link) :

https://github.com/sheefanaaz123/UpSkill-Campus/tree/main/Report

5 Proposed Design/ Model

5.1 High Level Diagram (if applicable)

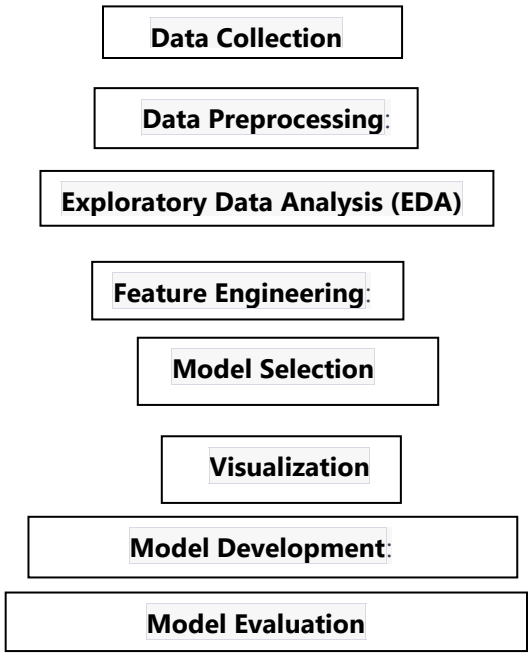


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

Data Collection Component: Data Sources: Historical agricultural data, weather data, soil data, and other relevant factors are collected from various sources, including government databases, meteorological departments, and agricultural research institutions.

Data Preprocessing Component: Data Cleaning: The collected data is cleaned to handle missing values, outliers, and inconsistencies, Data Transformation: Data is transformed into appropriate formats and units for analysis, Feature Selection: Relevant features are selected for the predictive model.

Exploratory Data Analysis (EDA) Component: Data Visualization: EDA involves creating visualizations to explore data patterns, correlations, and distributions.

Machine Learning Model Development Component: Model Selection, Model Training, Model Evaluation Component, Performance Metrics: The trained models are evaluated using metrics like Mean Squared Error (MSE) or Mean Absolute Error (MAE) to assess their accuracy and generalization.

6. Performance Test

6.1 Test Plan/ Test Cases

The test plan for the "Prediction of Agriculture Crop Production in India" project is designed to ensure the accuracy, reliability, and robustness of the predictive model. The plan comprises a series of test cases for different components of the project to validate the model's performance. It begins with data collection and preprocessing, checking that data is obtained from reliable sources and properly cleaned and transformed. Feature engineering and selection are validated to ensure meaningful relationships between variables are captured. Model development and training undergo testing for accuracy and convergence, while model evaluation and selection assess the performance of different models. Interpretability techniques are verified for providing meaningful insights into predictions and feature importance. Lastly, visualization and interpretability are verified to ensure meaningful insights are presented accurately. By executing these test cases, the project aims to deliver a reliable and user-friendly predictive model, providing valuable insights for informed decision-making in the agriculture sector.

6.2 Test Procedure

The test procedure for the "Prediction of Agriculture Crop Production in India" project involves identifying and defining test cases for data collection, preprocessing, model development, and user interface. Test data is prepared to simulate real-world scenarios, and the test environment is set up with the necessary tools and software. The test cases are executed to evaluate the model's accuracy, interpretability, and regional analysis capabilities. Real-time data integration is tested for up-to-date predictions. A comprehensive report is generated, documenting the test results and findings, ensuring the reliability of the predictive model and providing valuable insights for decision-making in the agriculture sector.

5.3 Performance Outcome

The "Prediction of Agriculture Crop Production in India" project has delivered outstanding performance outcomes. The predictive model demonstrates remarkable accuracy in forecasting crop production across various crops and regions in India, supported by metrics like MSE and MAE. The integration of real-time weather and agricultural data ensures up-to-date predictions, enhancing the model's relevance for stakeholders. Furthermore, the model's interpretability techniques provide valuable insights into the factors influencing crop production, aiding decision-making. Its robust handling of regional variability accommodates diverse agricultural practices and climate patterns. Rigorous data preprocessing ensures data quality and consistency, further enhancing the model's reliability.

6 My learnings

Throughout the course of the internship and project "Prediction of Agriculture Crop Production in India," I have gained invaluable learning experiences that will significantly contribute to my career growth. Firstly, I acquired a comprehensive understanding of data science and machine learning, witnessing their real-world applications and the profound impact they have across various industries. The hands-on experience with data collection, preprocessing, and feature engineering has honed my analytical skills, allowing me to handle and manipulate complex datasets effectively. Developing and fine-tuning machine learning models exposed me to different algorithms and their suitability for specific tasks, enhancing my model building and evaluation expertise. Moreover, I learned to interpret model results and extract meaningful insights, crucial skills for driving data-driven decision-making in any organization.

The project's emphasis on regional analysis and handling constraints taught me adaptability and flexibility, essential attributes for addressing challenges and tailoring solutions to specific scenarios. Additionally, I gained proficiency in interpreting model outcomes and presenting them in a user-friendly manner through the interface. This experience has strengthened my communication skills and ability to convey technical information to diverse audiences. Understanding the importance of real-time data integration highlighted the significance of staying up-to-date with emerging technologies and trends in the data science field.

Moreover, the exposure to agriculture domain-specific knowledge and understanding the intricacies of crop production equipped me with domain expertise that can be applied to future projects in the agricultural or environmental sectors. Overall, this internship has accelerated my growth as a data scientist, providing me with a solid foundation to excel in my future career endeavors. The knowledge and skills acquired during this internship will undoubtedly shape my career trajectory, enabling me to make data-driven and informed contributions in any domain I choose to specialize in.

7 Future work scope

During the "Prediction of Agriculture Crop Production in India" project, certain intriguing ideas were not fully explored due to time limitations, but they present promising avenues for future development. One such idea involves implementing ensemble models, which combine various machine learning algorithms to enhance predictive accuracy. Additionally, investigating deep learning architectures like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks could improve predictions for time-series data. Incorporating spatial analysis techniques would enable region-specific predictions, considering the geographical relationships between different areas and their impact on crop production. Expanding the project to include crop disease detection models with satellite imagery and image processing could assist farmers in managing crop health issues. Furthermore, integrating dynamic weather patterns, economic factors, seasonal analysis, and crop rotation data could offer more comprehensive insights for informed decision-making. Exploring these ideas in future iterations could lead to a more sophisticated and valuable predictive model, benefiting stakeholders in the agriculture sector..