Industrial Internship Report on

" Prediction of Agriculture Crop Production in India"

Prepared by

[Shivanjali & Yashvee Gupta]

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)
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

The project focuses on the comprehensive development of an AI-based fake news detection system. It aims to create a robust and versatile model capable of identifying fake news in various contexts. The core objectives of this project include evaluating detection performance using metrics like precision, recall, and F1-score, minimizing false positives to enhance user experience, and reducing false negatives to bolster threat detection.

To ensure the model's effectiveness, a diverse dataset encompassing fake and legitimate news is used, reflecting real-world scenarios. Real-time detection capabilities are also explored, aligning with the need for timely protection against misinformation.

Ethical considerations are a cornerstone, emphasizing the project's mission to combat fake news dissemination without inadvertently promoting it. Additionally, the project emphasizes educational value, aiming to raise user awareness about the dangers of fake news.

Comparisons with existing literature and works in the field provide insights into the project's contributions, advancements, and limitations. Areas for improvement include dataset expansion, feature engineering enhancement, model performance optimization, and minimizing false positives and negatives. Future extensions may include multilingual support, social media monitoring, image and video analysis, and integration into educational tools, enhancing the model's versatility and impact. This project embodies a holistic approach to addressing the pervasive issue of fake news in the digital age.

About need of relevant Internship in career development.

In today's rapidly evolving digital landscape, the need for data science internships has never been more relevant. Data science has become the backbone of virtually every industry, driving informed decision-making, enhancing operational efficiency, and enabling businesses to stay competitive. As organizations increasingly rely on data-driven insights, there is a pressing demand for skilled professionals who can harness the power of data. Data science internships provide a crucial bridge between theoretical knowledge and practical application. They offer aspiring data scientists the opportunity to gain hands-on experience in data collection, analysis, and interpretation. Moreover, these internships empower individuals to work with cutting-edge tools and technologies, such as machine learning and artificial intelligence, which are transforming industries from healthcare to finance. By offering real-world exposure and mentorship, data science internships not only meet the current demand for skilled professionals but also help shape the future of data-driven innovation. In a world where data is hailed as the new currency, data science internships are instrumental in preparing the next generation of talent to meet the evolving needs of industries and society.

Brief about Your project/problem statement.

Use Case Study: Prediction of Agriculture Crop Production in India

The use case study focuses on the critical task of predicting crop production in India, a country heavily reliant on agriculture. The objective is to develop a robust predictive model that can forecast crop yields for various crops and regions across the country. This project holds immense significance as accurate crop production forecasts can aid farmers, policymakers, and traders in making informed decisions, optimizing resource allocation, and ensuring food security.

The design of the use case study involves several key steps. First, we gather historical data on crop production, climatic conditions, soil properties, and other relevant factors from various sources, including government agencies, meteorological departments, and research institutions. Data preprocessing is then conducted to handle missing values, remove outliers, and standardize the data for analysis. Feature engineering plays a vital role in selecting the most influential variables affecting crop yields, such as temperature, rainfall, humidity, crop-specific characteristics, and geographical features.

With the prepared data, we delve into the realm of Machine Learning. The study involves exploring various regression and time series forecasting algorithms and evaluating their performances.

This use case study highlights the potential to create meaningful solutions that can positively impact India's agricultural sector. By harnessing the predictive capabilities of the developed model, stakeholders can make well-informed decisions to address challenges like food security, market fluctuations, and resource optimization, leading to sustainable growth in the agricultural landscape.

Opportunity given by USC/UCT.

Participating in a data science project or internship offers valuable opportunities for skill development and personal growth. These experiences provide hands-on exposure to essential technical skills like Python, R, data preprocessing, and machine learning, enabling individuals to become proficient in data handling and analysis. Additionally, interns gain expertise in statistical analysis and data visualization, enhancing their ability to draw meaningful insights from data. Problem-solving skills are honed as they tackle complex issues, while collaboration within interdisciplinary teams fosters teamwork and communication abilities. Moreover, domain-specific knowledge is acquired, making them valuable assets in various industries. These experiences also teach project management, ethical considerations, and effective communication. Successful completion of data science projects or internships not only enhances resumes but also provides a competitive edge in the job market, opening doors to rewarding careers in the dynamic field of data science.

How Program was planned
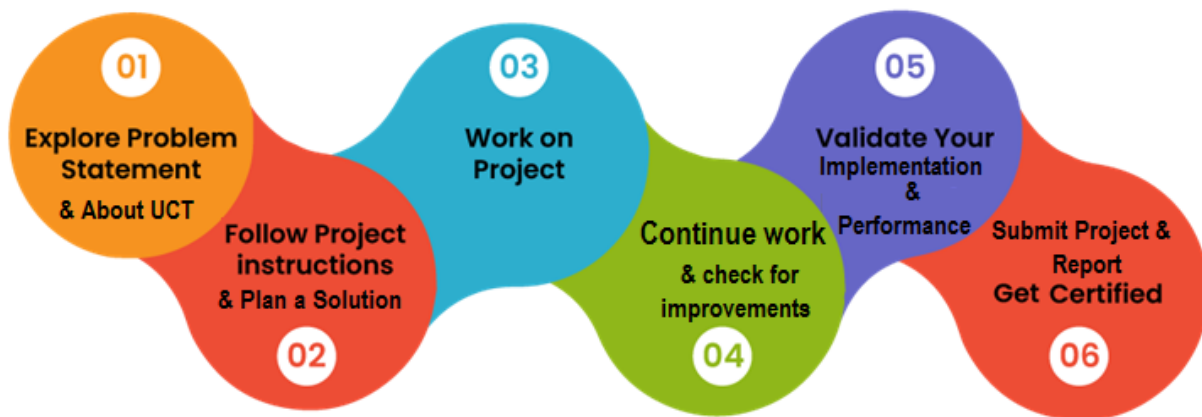
Your Learnings and overall experience

During my data science internship, I had the privilege of diving deep into the world of data analytics, and the learning experience was nothing short of transformative. From day one, I was exposed to the practical aspects of data preprocessing, which often involved wrangling messy datasets and dealing with missing values, a real-world challenge that academic courses often overlook. This internship also allowed me to develop a profound understanding of machine learning, as I had the opportunity to build predictive models, fine-tune hyperparameters, and evaluate model performance meticulously. The hands-on experience of working with industry-standard tools and libraries, such as Python, scikit-learn, and Pandas, significantly accelerated my technical skills and confidence in using these tools for real-world problem-solving.

Beyond the technical aspects, this internship taught me the importance of effective communication and collaboration. I had the privilege of working in a diverse and dynamic team, which challenged me to not only contribute to the project technically but also convey complex findings and insights in a clear and comprehensible manner. Presenting our results to stakeholders and discussing the practical implications of our analyses improved my ability to communicate data-driven insights effectively. Overall, my data science internship was a transformative journey that not only enhanced my technical competence but also fostered personal and professional growth, equipping me with valuable skills and experiences that will undoubtedly shape my future career in data science.

Thanks to all, who have helped directly or indirectly in this project.

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
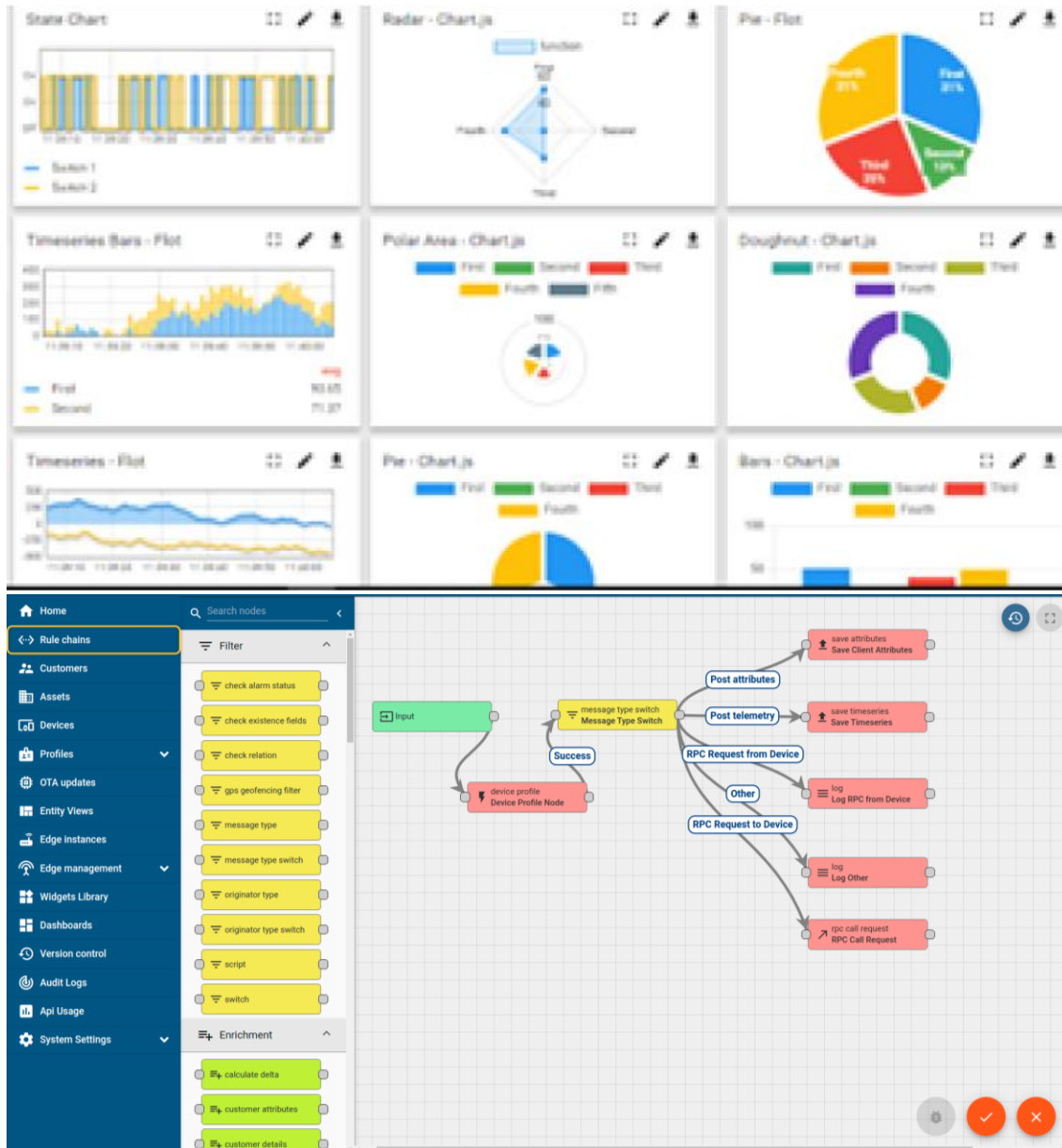
ii. **Smart Factory Platform (** FACTORY WATCH **)**

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 technology 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.
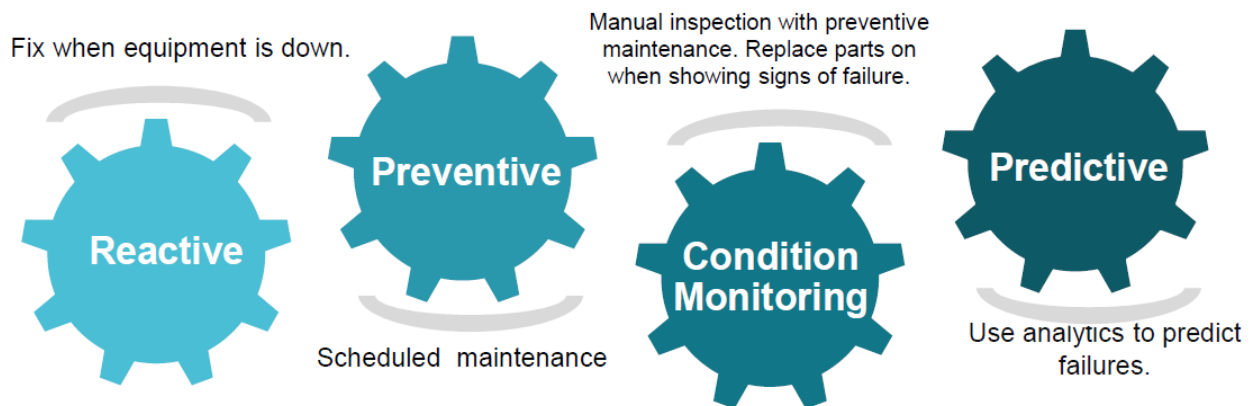
2.2   About upskill Campus (USC)

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.
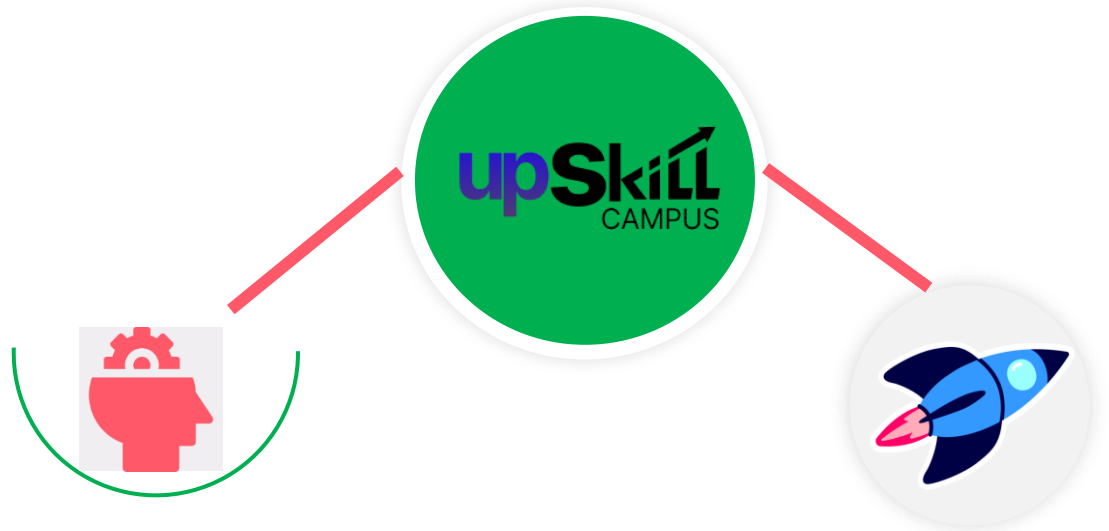
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/

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] Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network

approach International Journal of Innovative Research in Electrical, Electronics,

Instrumentation and Control Engineering vol 2 Issue 1 pp 683-6.

[2] Suresh A, Ganesh P and Ramalatha M 2018 Prediction of major crop yields of Tamilnadu using

K-means and Modified KNN 2018 3rd International Conference on Communication and

Electronics Systems (ICCES) pp 88-93 doi: 10.1109/CESYS.2018.8723956.

[3] Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning

techniques IEEE 5th International Conference for Convergence in Technology (I2CT) pp 1-5

doi: 10.1109/I2CT45611.2019.9033611.

3   Problem Statement

Prediction of Agriculture Crop Production in India

The use case study focuses on the critical task of predicting crop production in India, a country heavily reliant on agriculture. The objective is to develop a robust predictive model that can forecast crop yields for various crops and regions across the country. This project holds immense significance as accurate crop production forecasts can aid farmers, policymakers, and traders in making informed decisions, optimizing resource allocation, and ensuring food security.

The design of the use case study involves several key steps. First, we gather historical data on crop production, climatic conditions, soil properties, and other relevant factors from various sources, including government agencies, meteorological departments, and research institutions. Data preprocessing is then conducted to handle missing values, remove outliers, and standardize the data for analysis. Feature engineering plays a vital role in selecting the most influential variables affecting crop yields, such as temperature, rainfall, humidity, crop-specific characteristics, and geographical features.

With the prepared data, we delve into the realm of Machine Learning. The study involves exploring various regression and time series forecasting algorithms and evaluating their performances.

This use case study highlights the potential to create meaningful solutions that can positively impact India's agricultural sector. By harnessing the predictive capabilities of the developed model, stakeholders can make well-informed decisions to address challenges like food security, market fluctuations, and resource optimization, leading to sustainable growth in the agricultural landscape.

4   Existing and Proposed solution

The prediction of agricultural crop production in India has traditionally relied on historical data, manual observations, and expert knowledge. While this approach has been valuable over the years, it has limitations in terms of accuracy and scalability. The existing methods often lack the ability to analyze vast amounts of data, including satellite imagery, weather forecasts, and soil data, in real-time. These limitations can result in inaccurate predictions and challenges in adapting to changing climate conditions and emerging agricultural practice.

What is your proposed solution?

The proposed solution for predicting agricultural crop production in India revolves around a data-driven approach that harnesses the power of advanced technologies. By integrating diverse datasets encompassing historical yield records, satellite imagery, weather forecasts, soil quality information, and crop-specific data, we establish a robust foundation for predictive modeling. Employing cutting-edge machine learning techniques, including deep learning and ensemble methods, we aim to unearth intricate patterns and correlations within this extensive dataset. Real-time monitoring, facilitated by IoT sensors and remote sensing technologies, continuously collects vital information on factors like soil moisture, temperature, and crop health. This real-time data feeds into our predictive models, ensuring that crop yield forecasts remain up-to-date and adaptive to ever-changing conditions.

Climate modeling plays a pivotal role, enabling us to assess the impact of climate change on crop production. With this knowledge, both farmers and policymakers can take proactive measures to mitigate potential challenges. Our solution goes beyond prediction; it generates customized recommendations for farmers, offering guidance on optimal planting times, crop varieties, irrigation schedules, and strategies for pest control. To ensure accessibility and usability, we develop intuitive interfaces and mobile applications that provide real-time updates, weather alerts, and actionable insights to users. Scalability is a key consideration, as our solution is designed to accommodate the diverse agricultural practices and geographical variations found throughout India. Moreover, community engagement and farmer training programs are integrated, ensuring that our predictive insights translate into practical improvements in agricultural practices, fostering sustainability and food security for India's agricultural sector.

What value addition are you planning?

The proposed solution for predicting agricultural crop production in India offers significant value addition compared to traditional methods. By leveraging advanced data science and machine learning techniques, we bring several enhancements to the table.

Firstly, the integration of diverse datasets, including historical yield data, satellite imagery, weather forecasts, and soil information, enables a holistic view of the factors influencing crop production. This comprehensive data ecosystem forms the basis for more accurate and informed predictions.

Secondly, the adoption of cutting-edge machine learning models allows for the discovery of intricate patterns and relationships within the data, resulting in more precise crop yield forecasts. These models can adapt to changing conditions and provide real-time insights.

Thirdly, the inclusion of real-time monitoring through IoT sensors and remote sensing technologies ensures that the predictions remain current and relevant. This dynamic approach enables farmers to make timely decisions and adapt to evolving circumstances.

Moreover, the incorporation of climate modeling enhances the solution's capability to assess the impact of climate change on crop production, offering a proactive stance in addressing climate-related challenges.

Customized recommendations for farmers, based on predictive insights, empower them with actionable information, optimizing planting times, crop choices, and resource allocation.

User-friendly interfaces and mobile applications make these insights accessible to a wider audience, including farmers and stakeholders, enhancing the solution's practicality.

Finally, the emphasis on community engagement and farmer training programs ensures that the value generated by the solution translates into tangible improvements in agricultural practices, contributing to sustainability and food security in India.

4.1   Code submission (Github link)

https://github.com/Yashvee2007/upskillcampus/blob/main/PredictionOfAgricultureCropProductioninIndia.ipynb

4.2   Report submission (Github link) :

5   Performance Test

This is very important part and defines why this work is meant of Real industries, instead of being just academic project.

Here we need to first find the constraints.

How those constraints were taken care in your design?

What were test results around those constraints?

Constraints can be e.g. memory, MIPS (speed, operations per second), accuracy, durability, power consumption etc.

In case you could not test them, but still you should mention how identified constraints can impact your design, and what are recommendations to handle them.

###Random Forest Regressor

```
In [37]:   from sklearn.ensemble import RandomForestRegressor
           regr = RandomForestRegressor(max_depth=2, random_state=0, n_estimators=100)
           regr.fit(a_train, b_train)
           b_pred = regr.predict(a_test)

           from sklearn.metrics import mean_squared_error as mse
           from sklearn.metrics import mean_absolute_error as mae
           from sklearn.metrics import r2_score

           print('MSE =', mse(b_pred, b_test))
           print('MAE =', mae(b_pred, b_test))
           print('R2 Score =', r2_score(b_pred, b_test))

MSE = 7.67104887996405
MAE = 0.8953650873829122
R2 Score = 0.9589614680509004
```

XGBRegressor

```
In [39]: from xgboost import XGBRegressor
         from sklearn.metrics import mean_absolute_error
         XGBModel = XGBRegressor()
         XGBModel.fit(a_train,b_train , verbose=False)

         # Get the mean absolute error on the validation data :
         XGBpredictions = XGBModel.predict(a_test)
         MAE = mean_absolute_error(b_test , XGBpredictions)
         print('XGBoost validation MAE = ',MAE)
         XGBpredictions

         XGBoost validation MAE =  0.6670485576475581

Out[39]: array([1.0828081 , 0.64165634, 0.80906236, ..., 1.2118115 , 0.9264982 ,
                0.68652374], dtype=float32)

In [40]:
         print(r2_score(b_test , XGBpredictions))

         0.9654928330252374
```
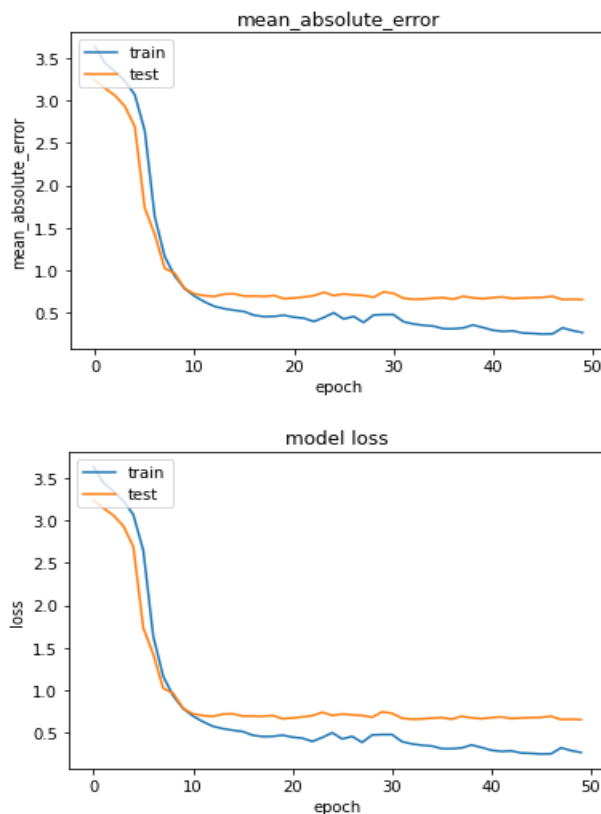
5.1 Test Plan/ Test Cases

5.2 Test Procedure

5.3 Performance Outcome

6 My learnings

Participating in this project has been an invaluable learning experience, encompassing both technical skills and broader insights. From a technical perspective, I have honed my abilities in data preprocessing, machine learning, and predictive modeling. Working with complex datasets and using advanced tools and libraries has significantly enhanced my proficiency in handling and analyzing data effectively.

Moreover, I've gained a profound understanding of the importance of interdisciplinary collaboration. In a team with diverse skills and perspectives, I've learned how to communicate complex technical findings in a clear and accessible manner to non-technical stakeholders, bridging the gap between data insights and actionable decisions.

The project has also underscored the real-world impact of data science. Witnessing how data-driven predictions can inform and optimize critical decisions, such as planting times for farmers, has reinforced the practical significance of this field.

Ethical considerations have been another crucial facet of my learning. Understanding the ethical implications of data collection, analysis, and dissemination is vital, and I've developed a heightened awareness of the importance of responsible data handling.

Furthermore, this project has instilled a sense of adaptability and resilience. Dealing with real-world data often means encountering unexpected challenges, and I've learned to approach problems with creativity and persistence, seeking innovative solutions.

7 Future work scope

The future scope of data science is exceptionally promising, with a multitude of exciting opportunities on the horizon. As artificial intelligence (AI) and machine learning (ML) continue to advance, data scientists will have the chance to create more sophisticated models and automated systems, revolutionizing industries across the board. The era of big data is here to stay, necessitating experts in data analytics to handle and derive insights from massive datasets. Additionally, the ethical dimension of data science will become increasingly crucial, with a growing demand for professionals who can develop AI models responsibly and mitigate biases. Data science's interdisciplinary applications will expand further, spanning fields like healthcare, finance, and environmental science, where data-driven insights will be instrumental in solving complex problems. Real-time data analysis, driven by the Internet of Things (IoT), will be a prevalent trend, demanding data scientists to design systems that can process and act on data in real-time, enhancing efficiency in various industries. As Natural Language Processing (NLP) and conversational AI evolve, data scientists will create more advanced chatbots, virtual assistants, and sentiment analysis tools. Quantum computing will usher in a new era of data analysis, unlocking unprecedented processing speeds. Edge computing will also rise in importance, requiring data scientists to develop models suitable for resource-constrained environments. The future of data science is not only technologically thrilling but also essential in addressing complex global challenges, ranging from healthcare to sustainability and beyond.