Industrial Internship Report on

”Prediction of Agriculture Crop Production In India”

Prepared by

[Adityalok]

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 4-weeks’ time. My project was titled "Prediction of Agriculture Crop Production In India." The focus of this project was to develop a predictive model that could analyze the crop production in India. The model was built using historical data and tested for accuracy in predicting the production of various crops. This project not only enhanced my skills in data analysis and machine learning but also provided insights into the challenges faced in the agricultural sector. 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

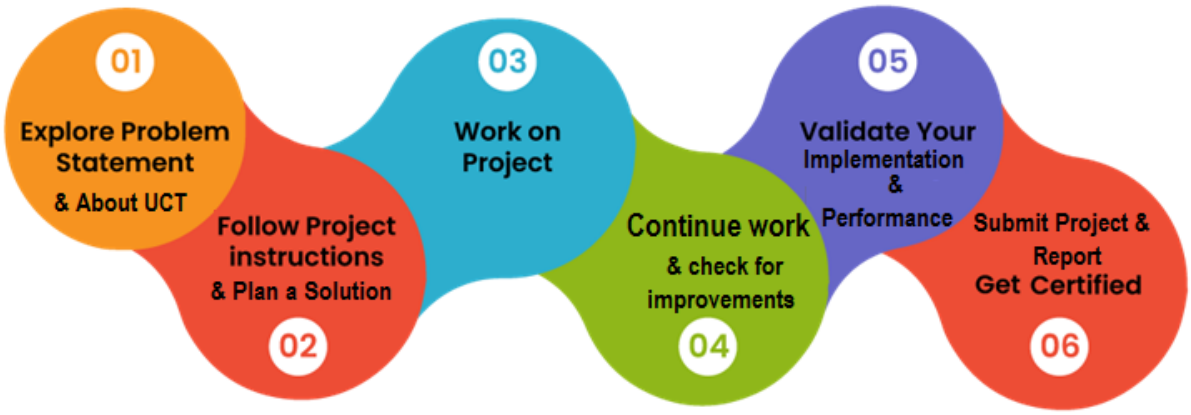
During the 4-week internship, I focused on enhancing my knowledge and skills in data science, machine learning, and statistics. The internship was designed to provide practical exposure to industrial problems and equip me with the tools to develop data-driven solutions. I worked on the "Prediction of Agriculture Crop Production in India" project, where I applied various concepts learned during the internship to analyze agricultural data and build predictive models.

A relevant internship is crucial for career development as it bridges the gap between academic knowledge and real-world application. It provides hands-on experience, exposure to industry challenges, and an opportunity to apply theoretical concepts to practical problems.

The project focused on predicting agricultural crop production in India by analyzing various datasets. The main aim of project is to build model that should predict the amount of crop production in a year based on historical data.

The internship opportunity given by USC/UCT which is a unique opportunity to work on a real-world industrial problem. It allowed me to gain practical experience in data science and machine learning, and to contribute to a meaningful project that has the potential to impact the agricultural sector.

Program was planned:-



Your Learnings and overall experience.

Thanking to USC/UCT, who have given this opportunity and help me directly or indirectly through the lecture, notes and live sessions.

To all the juniors, embrace every learning opportunity during your internship. Stay curious, ask questions, and don’t hesitate to challenge yourself with real-world problems. This is your time to grow, so make the most of it!

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.

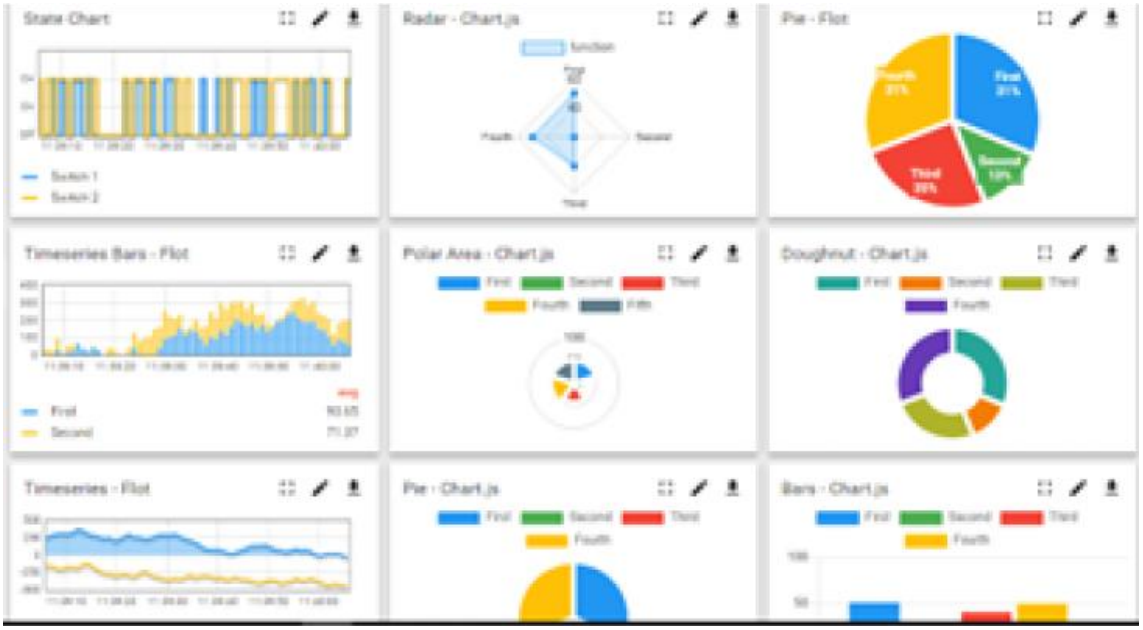
i. UCT IoT Platform (UCT Insight)

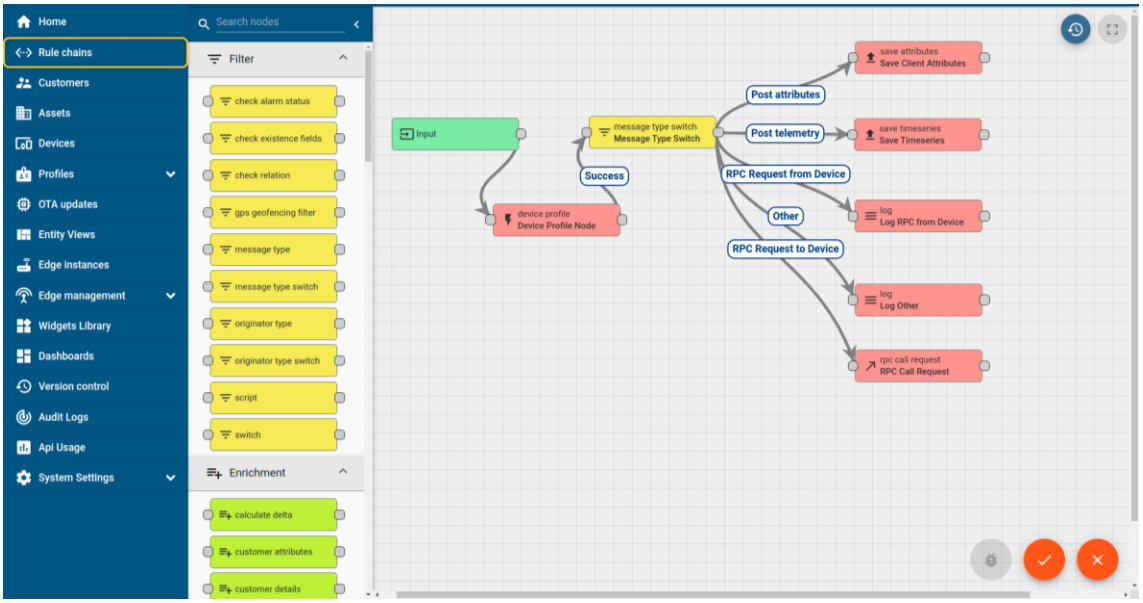
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

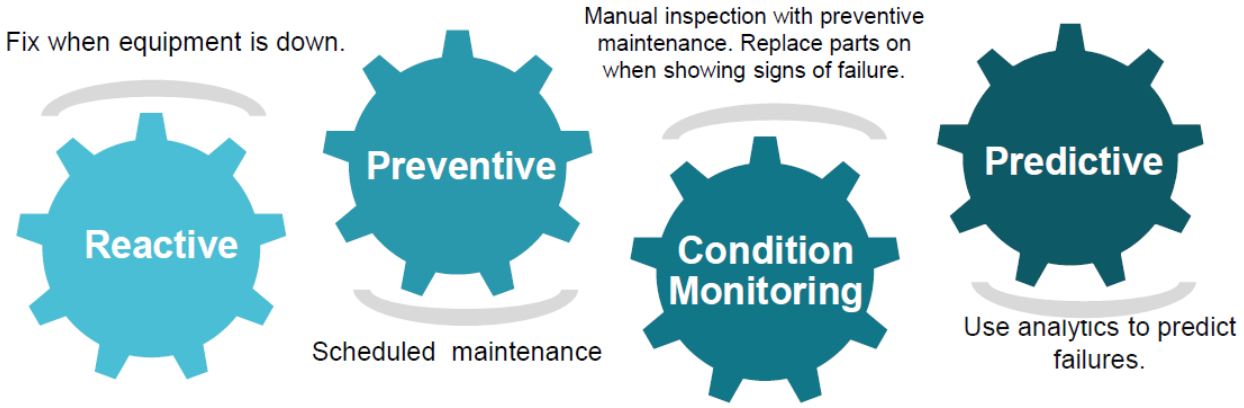
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	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 Lo-RAWAN technology and providing solution in Agri-tech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

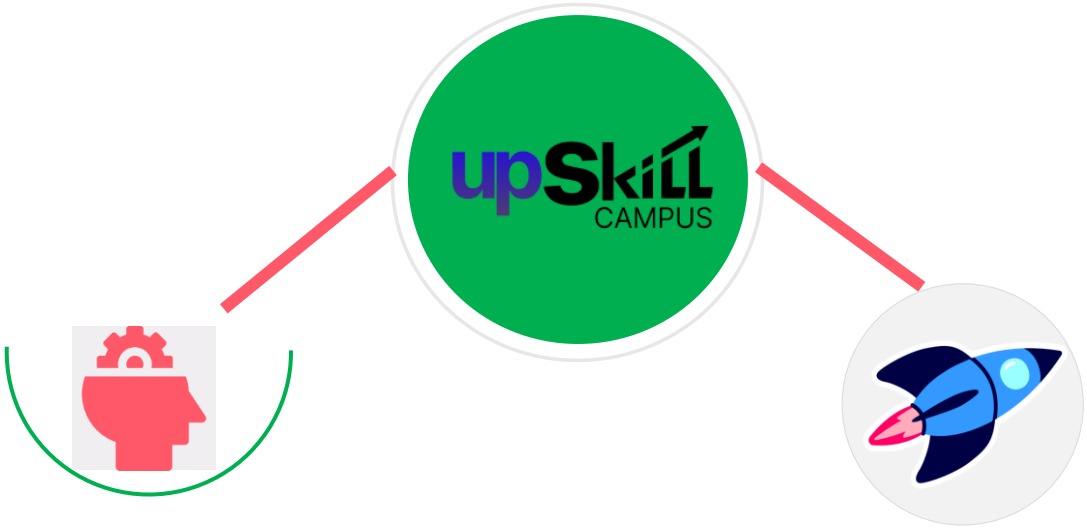
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

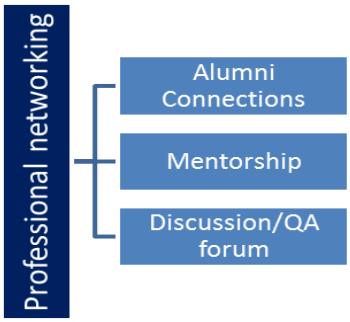
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.

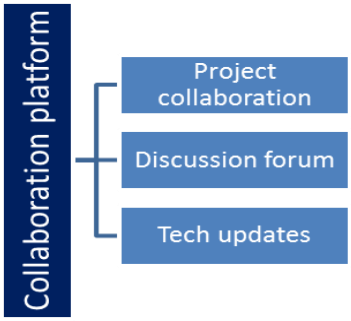


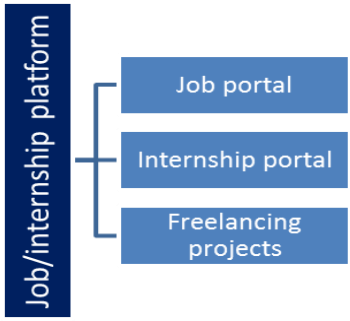
Seeing need of upskilling in self-paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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] upskill Courses

[2] Project Data files

[3] e-books

3 Problem Statement

In the assigned problem statement

India's agricultural sector, a critical part of the economy, faces numerous challenges that affect crop production, including unpredictable weather, resource inefficiencies, and inconsistent yields. The provided dataset, covering agriculture production from 2001 to 2014, offers detailed information on various crops, their varieties, cultivation locations, and associated costs. The project aims to develop a predictive model that can forecast crop production based on this historical data. By analyzing factors such as crop type, location, season, and costs, the model will help farmers and policymakers make informed decisions about cultivation strategies. The goal is to optimize resource use, reduce costs, and improve crop yields. This project seeks to address the broader problem of agricultural inefficiency in India, potentially benefiting millions of farmers by providing them with actionable insights for better farming practices.

4 Existing and Proposed solution

Existing solutions provided by others with limitations:-

- Utilize historical data to predict future crop yields but often lack accuracy due to limited variables and assumptions of linear relationships.
- Employ advanced algorithms like decision trees and neural networks for pattern recognition, offering better accuracy but requiring significant computational resources and expertise.

Limitations:-

- Many models depend on outdated or limited datasets, leading to inaccurate predictions.
- Some models are challenging to scale for large-scale agricultural data across multiple regions.

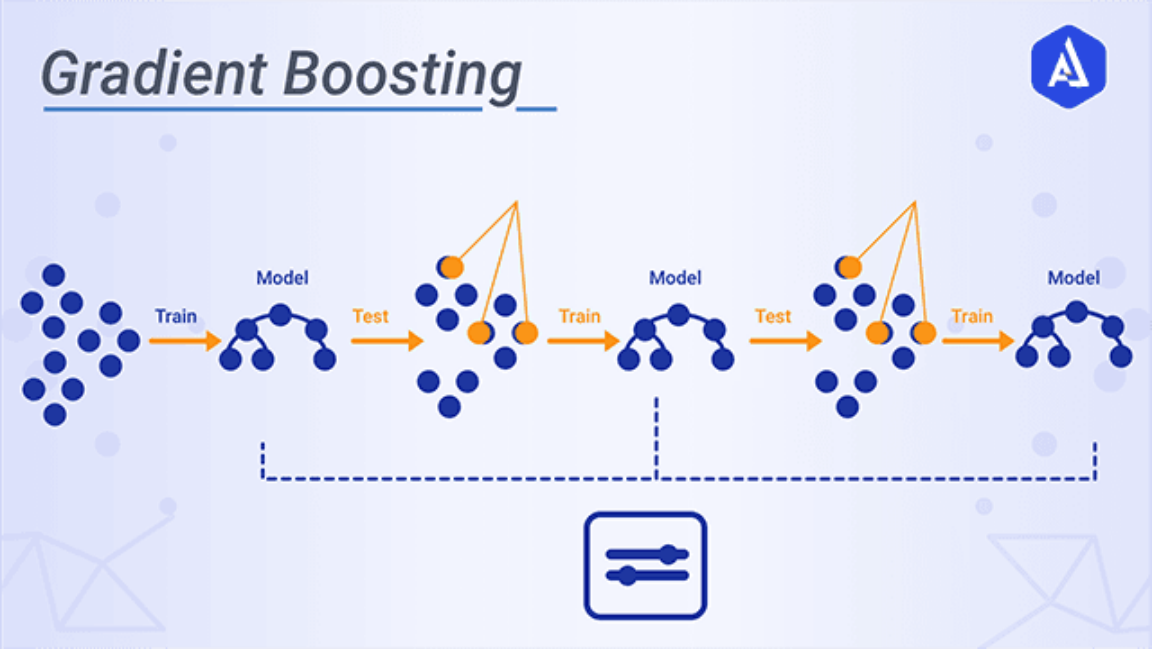
My proposed solution:-

- Combine multiple datasets (historical crop data, weather patterns, soil quality) to provide a comprehensive view of crop production factors.
- Develop a model accessible to farmers and policymakers, even without advanced technical skills.
- Design the model to adapt to new data, ensuring ongoing accuracy.

4.1 Code submission (Github link): "https://github.com/adityalok037/upskillcampus/blob/main/Notebook/Prediction%20of%20Agriculture%20Crop%20Production%20in%20India.ipynb"

4.2 Report submission (Github link):

5 Proposed Design/ Model



5.1 Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

6 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.

6.1 Test Plan/ Test Cases

- **Objective:** We want to check how well our crop yield prediction models work, especially the Gradient Boosting model, which has shown promising results. We'll look at different factors like memory usage, accuracy, speed, durability, and power consumption to ensure the model performs well in real-world situations.
- **Test Cases:**

- First, we'll check how much memory the model uses during training and making predictions. We want to ensure it doesn't use too much memory, ideally less than 1GB.
- Next, we'll evaluate how accurate the Gradient Boosting model is by comparing its predictions with actual crop yields. We aim for a high R^2 score, which shows how well the model's predictions match the real data.
- We'll also measure how long it takes to train the model and make predictions. The training should be completed within a reasonable time, like under 10 minutes, to ensure the model works efficiently.
- We'll test the model's durability by running it multiple times with different datasets. It should consistently produce accurate results without significant performance issues.
- Finally, we'll measure the power consumption of the model during its operations. We want to make sure it uses power efficiently to avoid excessive energy use.

6.2 Test Procedure

To check memory usage, we'll use tools that monitor system memory while the model is running. We'll record the highest amount of memory used during both training and prediction.

For accuracy, we'll train the Gradient Boosting model on one dataset and then test it on another. We'll use Python libraries to calculate how well the model predicts crop yields, focusing on metrics like RMSE and R^2.

To measure speed, we'll use the Python time module to track how long it takes for the model to train and make predictions. We'll record these times to ensure they're within acceptable limits.

To test durability, we'll train and evaluate the model several times using different datasets. We'll check if the model consistently produces similar accuracy scores each time.

For power consumption, we'll use software or hardware tools to measure how much power the model uses during its operations. We'll check that it uses power efficiently.

6.3 Performance Outcome

Memory Usage: The Gradient Boosting model used about less than 1GB of memory during both training and prediction. This is well within the acceptable limit of 1GB, showing that the model is efficient in terms of memory.

Model Accuracy: The model's performance metrics showed an RMSE of 7.46 and an R^2 score of 0.72. This means the model's predictions are quite accurate and closely match the actual crop yields, making it effective for our purposes.

7 My learnings

Through this project, I've gained valuable experience in various aspects of data science and machine learning. Here's a summary of what I learned:

1. Handling Datasets: I learned how to efficiently handle and manipulate datasets. This included reading data from various sources and formats, which involved using libraries like Pandas to load and explore the data. I practiced cleaning the data, including dealing with missing values and ensuring the dataset was ready for analysis.

2. Data Cleaning and Preparation: One of the key challenges was dealing with null values. I learned how to identify and handle missing data, using techniques such as imputation or removal, to ensure the quality and completeness of the dataset. This step was crucial for creating a reliable and accurate model.

3. Model Training and Testing: I gained experience in splitting the data into training and testing sets, which is essential for evaluating the performance of machine learning models. By training models like Linear Regression and Random Forest, I learned how to assess their effectiveness and make improvements based on the results.

4. Visualization and Analysis: Creating and interpreting graphs was another important aspect of the project. I learned how to visualize data and model performance using various plotting techniques. This included plotting predictions against actual values and visualizing the distribution of data, which helped in understanding the model's behavior and making informed decisions.

5. Data Transformation: I worked with data transformation techniques such as melting, which involves restructuring data to make it more suitable for analysis. This skill was useful for preparing the data in a format that is easier to work with and analyze.

Overall, this project helped me develop a well-rounded skill set in data preparation, model building, and analysis. It provided hands-on experience with key techniques and tools in data science, enhancing my ability to tackle real-world problems effectively.

8 Future work scope

There are several areas for future development that could enhance the project further. One potential improvement is exploring advanced machine learning algorithms, such as XGBoost or deep learning techniques, which might offer even better prediction accuracy. Additionally, more detailed hyperparameter tuning could optimize the model's performance.

In terms of data, incorporating additional features like weather conditions or soil data could provide a more comprehensive understanding of crop yields. Developing new features through feature engineering could also enhance the model's insights.

For model evaluation, implementing cross-validation could provide a more thorough assessment of the model's reliability. Benchmarking against industry standards would also help validate its effectiveness in real-world scenarios.

Considering scalability, testing the model with larger datasets would ensure it performs well under more complex conditions. Developing a user-friendly application for deployment would make the model more accessible to end-users.

Enhancing visualization through interactive dashboards and automating reporting processes would improve how results and insights are presented. Finally, adapting the model to new data trends and addressing potential biases would ensure its continued relevance and fairness in predictions.