Industrial Internship Report on

” Advancing Crop Production in India through Predictive Analytics”

Prepared by

Harshala B

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy incollaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a Quality Prediction in mining process project provided by UCT. We had tofinish the project including the report in 6 weeks’ time. My project was to explore real industrial data and help manufacturing plants to be more efficient The main goal is to use this data to predict how much impurity is in the ore concentrate. As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we canhelp the engineers, giving them early information to take actions. Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the number of ore that goes to tailings as you reduce silica in the ore concentrate). 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

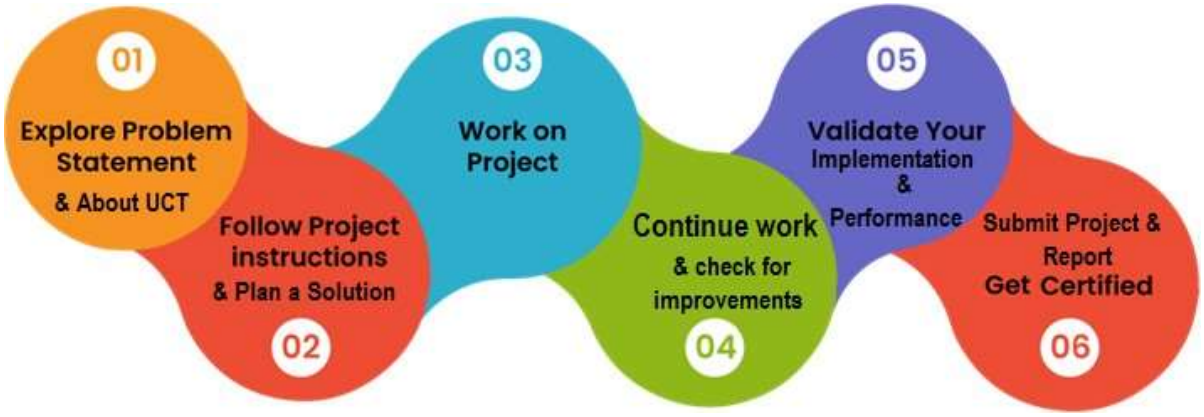
1 Preface

In the 6-week Internship program we were able to understand the concepts of Data Science and Machine Learning.

- Data Science: Introduction to the Data Science and we were able to gain the knowledge about the big data which is also created by the universe. The applications of data science in different domain areas and the usage of modern technology in our daily life.
- Machine learning: The origin of Machine learning and its development with the definition and its application were explained. Relation of the Artificial Intelligence, Machine Learning and Deep learning concepts knowledge. Explanation of the machine learning working with the block diagram helped us to explore our project.

1.1.1 Abstract

The project, "Advancing Crop Production in India through Predictive Analytics," aims to transform Indian agriculture by harnessing data science and machine learning. It begins by collecting and integrating diverse datasets, including historical weather patterns, soil quality, and crop yields from across India. These data sources are consolidated into a comprehensive database. The project's central focus is on developing advanced predictive models using state-of-the-art machine learning techniques. These models provide valuable insights such as precise crop yield forecasts, early warnings for potential pest outbreaks, and recommendations for optimal planting times. Real-time inputs from weather stations and satellite imagery are incorporated for up-to-date accuracy. The potential impact of this project is significant. It enables precision agriculture, offering farmers actionable guidance for resource-efficient decision-making. For instance, farmers will receive specific recommendations on when and what to plant, ultimately reducing resource wastage and increasing crop yields. Furthermore, the project addresses agricultural risk mitigation. By providing early warnings for pests and adverse weather events, it empowers farmers to take proactive measures to protect their crops, minimizing losses and enhancing food security. From a policy perspective, this project contributes data-driven insights for policymakers. Comprehensive agricultural trend analyses enable the creation of more effective policies that support sustainable and profitable farming practices. Economically, the implications are substantial. Enhanced agricultural productivity not only bolsters food security but also drives economic growth. Increased crop yields translate into higher income opportunities for farmers, reduced reliance on food imports, and a more resilient national economy. In summary, "Advancing Crop Production in India through Predictive Analytics" represents a pioneering fusion of data science, machine learning, and agriculture. Its overarching goal is to make Indian agriculture more sustainable, resilient, and profitable by leveraging predictive analytics. This promises a brighter future for both farmers and the nation as a whole.


01
Explore Problem Statement
& About UCT
02
Follow Project instructions
& Plan a Solution
03
Work on Project
04
Continue work
& check for improvements
05
Validate Your Implementation & Performance
06
Submit Project & Report
Get Certified

Learning Highlights:

i. Gained detail information about Data science and machine learning.

ii. Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more.

iii. Concepts of model training and model Evaluation in machine learning as trained model, test model and validation model.

iv. AI technologies more demanding significance.

v. A good understanding of big data platforms like Hadoop, Statistical analysis is improved for this path.

I want to Thank Mr. Nithin Tyagi and Mr. Ankit for being a wonderful support network, your guidance and shared experiences have been invaluable. I appreciate everything you both have intimated the instructions and guidelines to complete my internship successfully.

I thoroughly enjoyed my internship this summer and now have very valuable experience under my belt. I know this will help when looking for jobs and needing references. Practical experience is the best and internships give students that hands on experience they need. I feel that quality internships are essential to develop key skills that we can't get in a classroom. Upskill campus provided us the practical experience and knowledge about the Data Science and Machine learning, helped to select the projects and completion of it. Thank you Upskill, campus, for providing me this internship opportunity.

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.

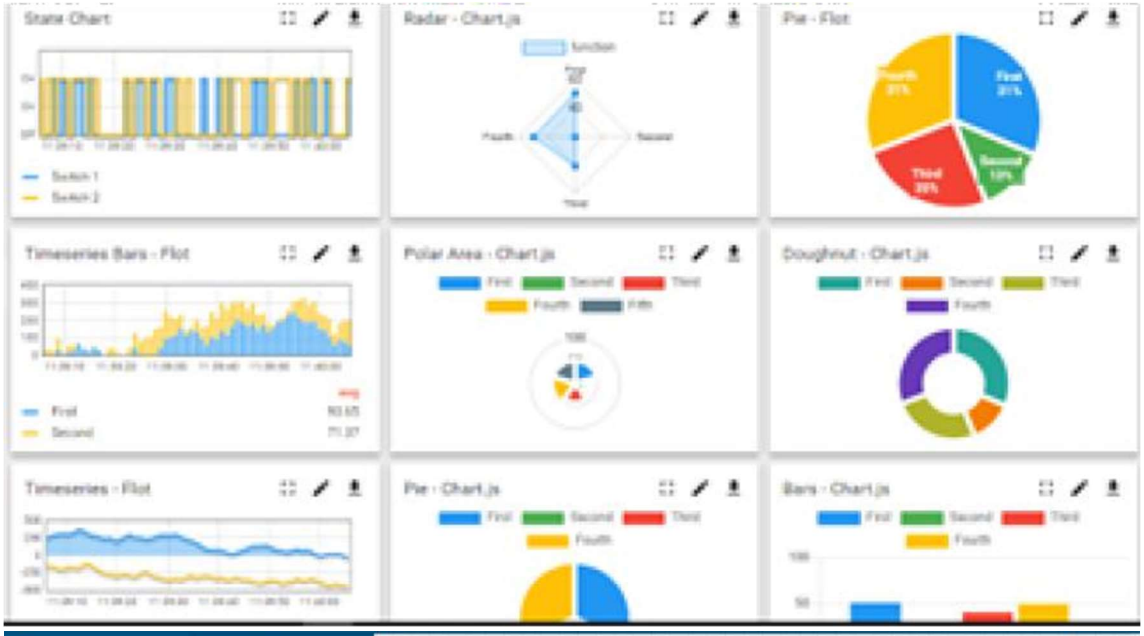
i. UCT IoT Platform (uct Insight)

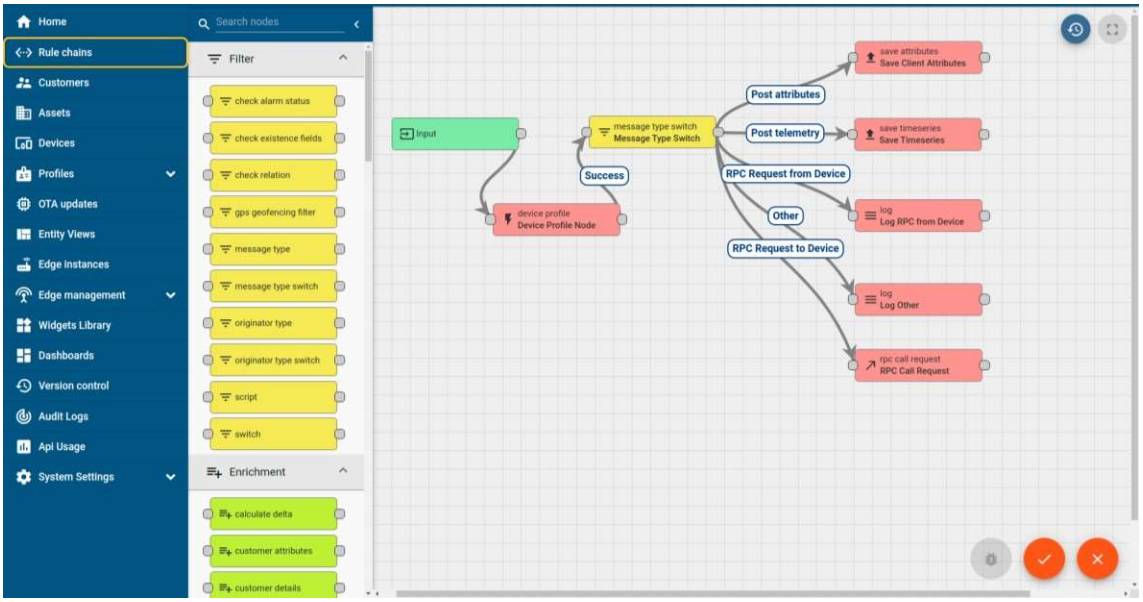
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

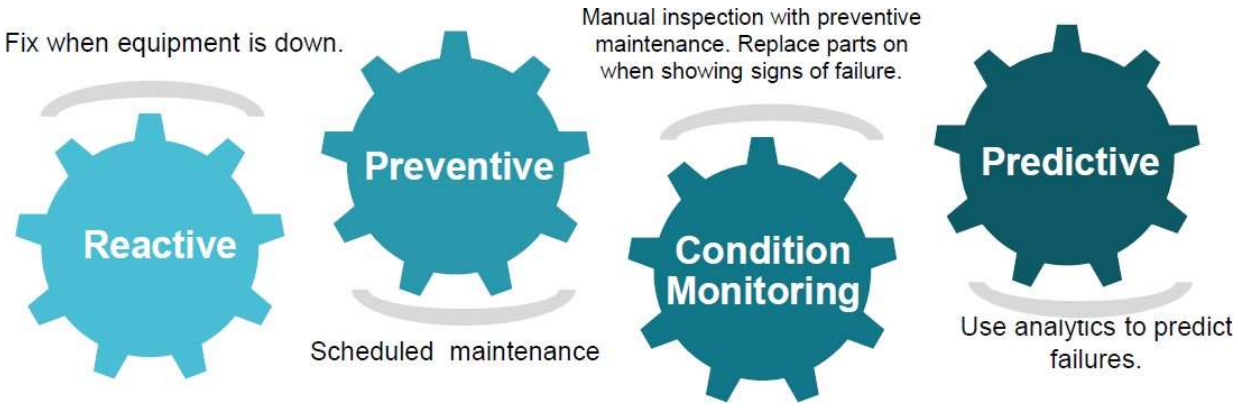
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

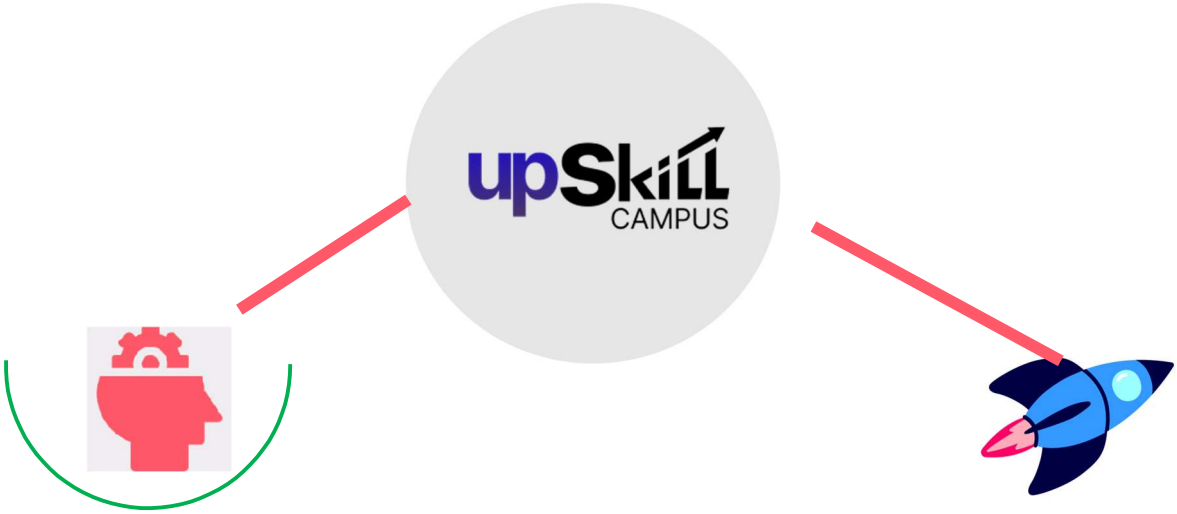
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

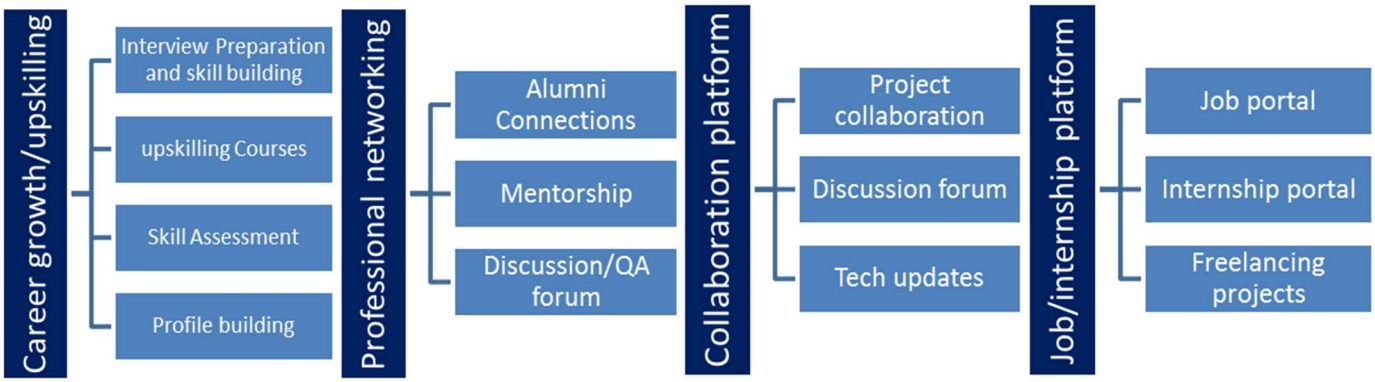
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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 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.

2.4 Reference

[1] https://www.jagranjosh.com/general-knowledge/list-of-major-crops-of-india-temperature-rainfall-soil-1473918924-1

[2] https://www.kaggle.com/code/theeyeschico/crop-analysis-and-prediction

[3] https://www.researchgate.net/publication/323204575_Crop_prediction_using_predictive_analytics

[4] https://www.sciencedirect.com/science/article/pii/S1877050920310449

[5] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00668-2

[6] https://www.mdpi.com/2076-3417/13/16/9288

[7] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9557312

3 Problem Statement:

Utilizing data science and machine learning techniques to forecast agricultural crop production in India, to develop predictive models that can accurately estimate crop yields based on factors such as climate data, soil characteristics, historical production records, and other relevant variables.

3.1 AIM: To utilize data-driven insights and predictive models to forecast agricultural crop production in the context of India.

The goal is to develop a comprehensive understanding of various factors influencing crop yields and create predictive models that can assist in making informed decisions for the agricultural sector. The project's significance lies in its potential to revolutionize agricultural practices in India. By accurately predicting crop production, farmers, policymakers, and agricultural businesses can make more informed decisions, optimize resource utilization, and mitigate risks associated with unpredictable crop yields. Additionally, this project contributes to the advancement of data science applications in real-world scenarios, showcasing the power of technology in addressing complex challenges.

3.2 Objective:

I. Data Collection: Gather relevant data sets that encompass historical crop production records, climate data, soil characteristics, and other pertinent variables that contribute to crop yields

II. Exploratory Data Analysis (EDA): Perform an in-depth analysis of the data to identify patterns, correlations, and trends. EDA helps in gaining insights into the relationships between various factors and crop production

III. Model Development: Utilize machine learning algorithms to develop predictive models. These models should be capable of forecasting crop production based on inputs such as weather conditions, soil properties, and historical production data.

IV. Model Evaluation: Assess the performance of the predictive models using appropriate metrics. This step ensures that the models provide reliable predictions and can be trusted for decision-making

V. Scenario Analysis: Apply the developed models to different scenarios, such as varying weather conditions or changes in agricultural practices. This can provide valuable insights into potential outcomes and inform planning.

4 Existing and Proposed solution:

Existing Solution:
Currently, there are several ongoing efforts and existing solutions aimed at advancing crop production in India through predictive analytics:

Weather Forecasting: Weather data is crucial for farmers to make informed decisions. The India Meteorological Department (IMD) provides weather forecasts, which can be integrated with predictive analytics tools to help farmers plan their planting and harvesting schedules.

Satellite Imagery: Organizations like ISRO (Indian Space Research Organization provide satellite imagery that can be used for monitoring crop health, identifying disease outbreaks, and assessing soil moisture levels.

Crop Health Monitoring: Startups and research institutions have developed remote sensing technologies and mobile apps that enable farmers to monitor the health of their crops using drones and smartphones. These tools often employ machine learning algorithms to detect diseases and pests early.

Soil Health Analysis: Predictive analytics can be used to analyze soil data, providing insights into nutrient levels, pH, and salinity. Organizations like the National Soil Data Repository and various state agricultural departments collect and disseminate soil information

Proposed Solution:
To further advance crop production in India through predictive analytics, a comprehensive solution can be developed and expanded upon:

Integrated Data Platform: Create a centralized platform that aggregates data from various sources, including weather forecasts, satellite imagery, soil health data, and historical crop performance. This platform should be accessible to farmers, agricultural experts, and policymakers.

Machine Learning and AI: Develop machine learning models and artificial intelligence algorithms that can process the integrated data and provide predictive analytics. These models can forecast crop yields, identify disease outbreaks, recommend optimal planting dates, and suggest irrigation schedules.

Mobile Applications: Develop user-friendly mobile applications that farmers can easily access to receive personalized recommendations based on predictive analytics. These apps should consider local conditions and provide actionable insights in regional languages.

Training and Education: Implement training programs to educate farmers about the benefits of predictive analytics and how to use the provided tools effectively. This includes both digital literacy and agricultural best practices.

Government Support: Collaborate with government agencies to support the adoption of predictive analytics in agriculture. This could involve subsidies for technology adoption, incentivizing data sharing, and creating policies that encourage sustainable farming practices.

Research and Development: Continue investing in research to improve predictive models and data collection methods. Collaborate with agricultural universities and research institutions to stay at the forefront of agricultural technology.

Data Security and Privacy: Ensure robust data security and privacy measures to protect sensitive information. Farmers must feel confident that their data will not be misused.

Scaling and Accessibility: Focus on making predictive analytics tools accessible to smallholder farmers, who make up a significant portion of India's agricultural workforce. Consider partnerships with microfinance institutions to facilitate access.

Feedback Mechanism: Establish a feedback loop for farmers to report issues, share their experiences, and suggest improvements. Continuous improvement is key to the success of any agricultural technology initiative.

By implementing these proposed solutions, India can harness the power of predictive analytics to enhance crop production, improve food security, reduce production costs, and promote sustainable agriculture practices. This integrated approach can benefit both farmers and the entire agricultural ecosystem.

4.1 Code submission (GitHub link)

upskillcampus/AdvancingCropProductioninIndia.ipynb at main · HarshalaaB/upskillcampus (github.com)
Code File name: AdvancingCropAnalysisinIndia.ipynb

4.2 Report submission (GitHub link)

HarshalaaB/upskillcampus (github.com)

Report File name: AdvancingCropProductioninIndia_Harshala_B_USC_UCT.pdf

5 Proposed Design/ Model

Designing a model for advancing crop production in India through predictive analytics involves creating a systematic framework that integrates data, technology, and processes to provide actionable insights for farmers and agricultural stakeholders.

1.Data Integration:

- Data Sources: Identify and collect data from various sources, including weather agencies, satellite imagery providers, soil health repositories, and historical crop data.
- Data Preprocessing: Clean, preprocess, and format the data to make it suitable for analysis. This may involve handling missing values, outliers, and data normalization.

2. Centralized Data Platform:

- Data Warehousing: Store the integrated data in a centralized data warehouse or cloud-based storage system for easy access and scalability.
- Real-time Updates: Ensure that data is updated regularly, especially weather and satellite data, to provide real-time insights.

3. Predictive Analytics:

- Machine Learning Models: Develop predictive models using machine learning algorithms. Some common models include regression, decision trees, and neural networks.
- Feature Engineering: Extract relevant features from the data, such as temperature, humidity, rainfall, crop type, and soil health indicators.
- Model Training: Train the machine learning models using historical data to predict outcomes like crop yields, disease outbreaks, and optimal planting/harvesting times.
- Validation: Validate the model's accuracy and performance through cross-validation and testing on independent datasets.

4. User-Friendly Interface:

- Mobile and Web Apps: Create user-friendly applications accessible on smartphones and web browsers. Consider a multilingual interface to accommodate users from diverse linguistic backgrounds.
- Dashboard: Provide a dashboard with personalized recommendations and visualizations, including weather forecasts, crop health status, and soil conditions.

5. Personalized Recommendations:

- AI-Driven Insights: Use AI algorithms to generate personalized recommendations for farmers, taking into account their location, crop type, and historical data.
- Actionable Advice: Offer actionable advice, such as suggesting irrigation schedules, pest management strategies, and crop rotation plans.

6. Training and Support:

- Farmer Training: Develop training programs and resources to educate farmers on how to use the platform effectively and interpret predictive insights.
- Customer Support: Establish a customer support system for farmers to seek assistance and troubleshoot issues.

7. Government and Institutional Integration:

- Collaboration: Partner with government agricultural departments, research institutions, and NGOs to ensure widespread adoption and support.
- Policy Support: Advocate for policies that promote the use of predictive analytics in agriculture, including subsidies for technology adoption.

8. Data Security and Privacy:

- Secure Data Handling: Implement robust security measures to protect sensitive agricultural data and ensure compliance with data privacy regulations.
- Data Ownership: Clearly define data ownership and usage rights to build trust among stakeholders.

9. Continuous Improvement:

- Feedback Loop: Establish a mechanism for users to provide feedback and report issues, enabling continuous improvement of the platform.
- Research and Development: Invest in ongoing research to enhance predictive models and incorporate emerging technologies.

10. Scalability and Accessibility:

- Scalable Architecture: Design the platform to handle a growing user base and increasing data volumes.
- Accessibility for All: Ensure that the platform is accessible to smallholder farmers and those in remote areas, possibly through partnerships with local organizations.

11. Impact Assessment:

- Monitoring and Evaluation: Continuously monitor the impact of the predictive analytics platform on crop yields, farmer income, and food security.
- Adjustment: Use feedback and impact assessments to make necessary adjustments and refinements to the model

Machine Learning

The steady progress of machine learning has been quite phenomenon over a decade where data is now seen as an important asset to be used judiciously by all companies. Machine learning, in simple term referred to algorithms that learned from data in an iterative manner to identify trends, patterns and correlations. Moreover, it is more applicable especially to datasets where past observations are potent predictor of the future. In this study, among wide variety of existing methods, we employed three supervised machine learning algorithms to a historical dataset observed in the froth floatation plant system.

Supervised machine learning is a process of furnishing the algorithm with observations in which the variable outcome of interest is known beforehand and the algorithm learn from the observed data to make prediction of future values.

6 Performance Test

1.Objective:

- To evaluate the performance, scalability, and reliability of the predictive analytics system for crop production in India.

2. Performance Metrics:

- Response Time: Measure the time taken to generate predictions, recommendations, and reports.
- Throughput: Assess the system's ability to handle a certain number of requests per unit of time.
- Resource Utilization: Monitor CPU, memory, and disk usage during peak loads.
- Scalability: Determine how the system scales as the workload increases.
- Reliability: Evaluate the system's uptime, error rates, and fault tolerance.
- Concurrency: Measure how well the system handles multiple simultaneous users or requests.

3. Test Scenarios:

- Baseline Test: Measure the system's performance under normal conditions.
- Load Testing: Gradually increase the load (simulated users/requests) to assess system response and identify breaking points.
- Stress Testing: Apply extreme load conditions to determine the system's behavior under duress.
- Scalability Testing: Assess how well the system scales by adding more resources or nodes.
- Failover and Recovery Testing: Test the system's ability to recover from failures gracefully.
- Security Testing: Evaluate the system's resistance to security threats and attacks.
- Peak Load Testing: Assess performance during peak usage periods, such as planting and harvesting seasons.
- Geographic Load Testing: Simulate usage from various geographic locations to ensure global accessibility.

6.1.1 Test Plan/ Test Cases

Objectives: The test plan outlines the objectives of testing, which include verifying the accuracy of the machine learning model in predicting silica and iron concentrations in mining samples.

1. Test Data:

- Use realistic datasets that represent the diverse agricultural conditions in India, including different crops, regions, and weather patterns.

2. Test Execution:

- Execute each test scenario using automated testing tools to simulate user interactions and generate load.

- Continuously monitor system performance and collect relevant metrics.

3. Performance Acceptance Criteria:

- Define acceptable performance thresholds for each metric (e.g., response time < 3 seconds).
- Identify performance bottlenecks and areas that need improvement.

4. Reporting:

- Generate comprehensive performance reports, including test results, performance metrics, and recommendations for improvement.

5. Iterative Improvement:

- Based on test results, iteratively improve the system to meet or exceed performance expectations

6.1.2 Test Procedure

- Data Collection: Gather relevant data sets that encompass historical crop production records, climate data, soil characteristics, and other pertinent variables that contribute to crop yields
- Test Execution: Executing each test case by loading input data, running the model, and comparing predictions to expected values.
- Data Validation: Calculating evaluation metrics (e.g., MAE, MSE) to measure prediction accuracy.
- Post-Conditions: Recording results, including predicted values and evaluation metrics.

6.1.3 Performance Outcome

1.Increased Crop Yield:

- Measure the percentage increase in crop yields compared to previous years or to a control group that does not use predictive analytics.
- Assess the average increase in yield across different crops and regions.

2.Reduced Crop Losses:

- Calculate the reduction in crop losses due to factors such as weather-related events, pests, and diseases.
- Evaluate the economic value saved by reducing losses.

3.Improved Resource Efficiency:

- Measure the reduction in resource usage, such as water and pesticides, per unit of crop produced.
- Assess the impact on resource conservation and cost savings.

4.Farmers' Income Growth:

- Evaluate the increase in farmers' income resulting from higher yields and reduced losses.
- Assess the overall economic impact on rural communities.

Output of the data:

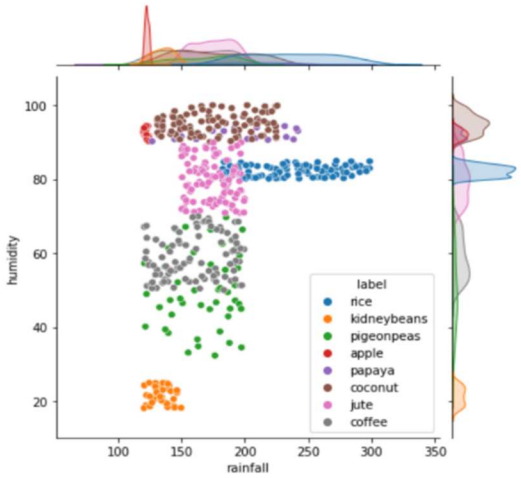


Fig. Rainfall and Humidity

During rainy season, average rainfall is high (average 120 mm) and temperature is mildly chill.
Rain affects soil moisture which affects ph of the soil. Here are the crops which are likely to be planted during this season.

- Rice needs heavy rainfall (>200 mm) and a humidity above 80%.
- Coconut is a tropical crop and needs high humidity therefore explaining massive exports from coastal areas around the country.

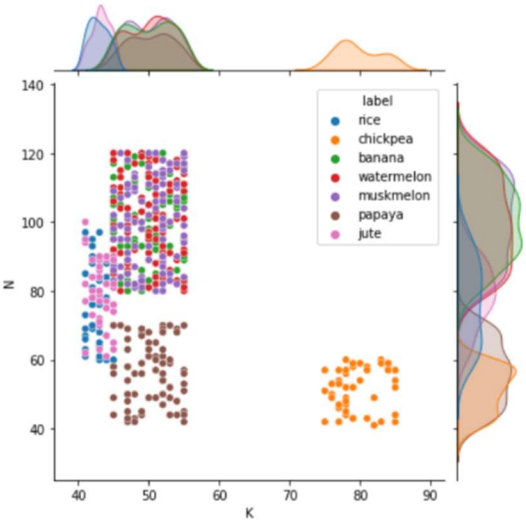


Fig. Average Potassium and nitrogen concentration

- This graph correlates with average potassium (K) and average nitrogen (N) value (both>50).
- These soil ingredients directly affect nutrition value of the food. Fruits which have high nutrients typically has

consistent potassium values.

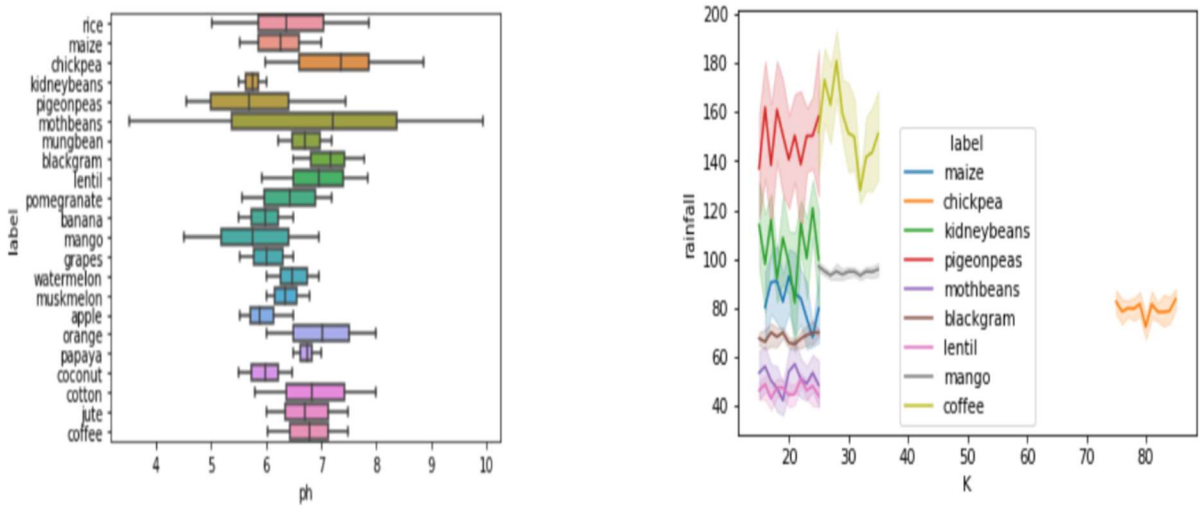


Fig. Phosphorous requirement for different crops

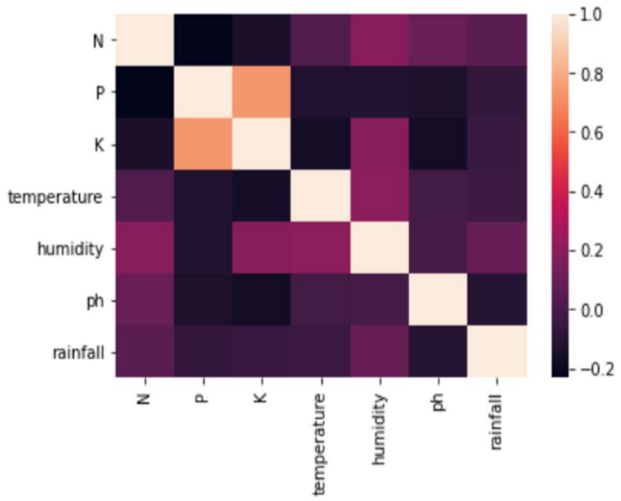


Fig. Correlation visualization between features

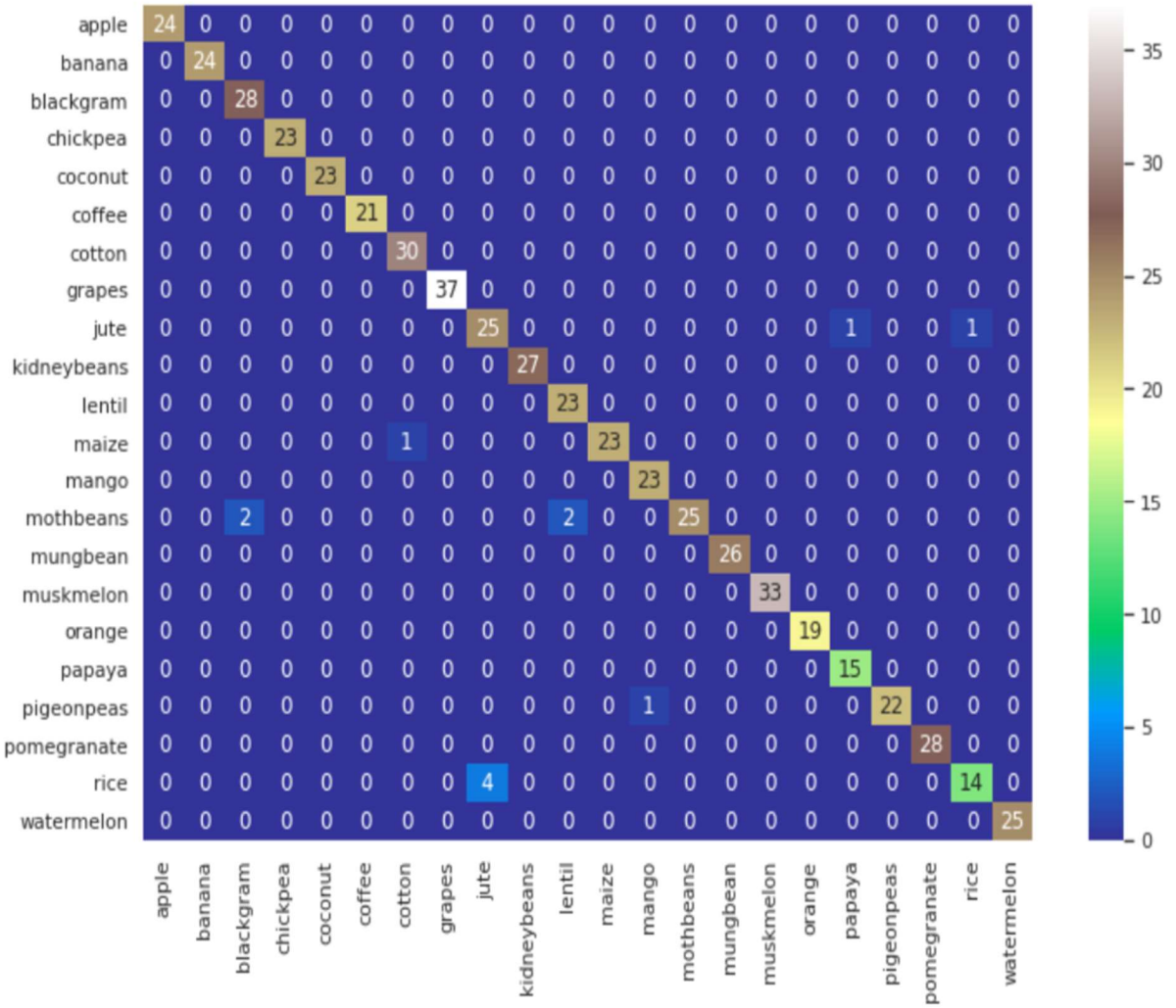
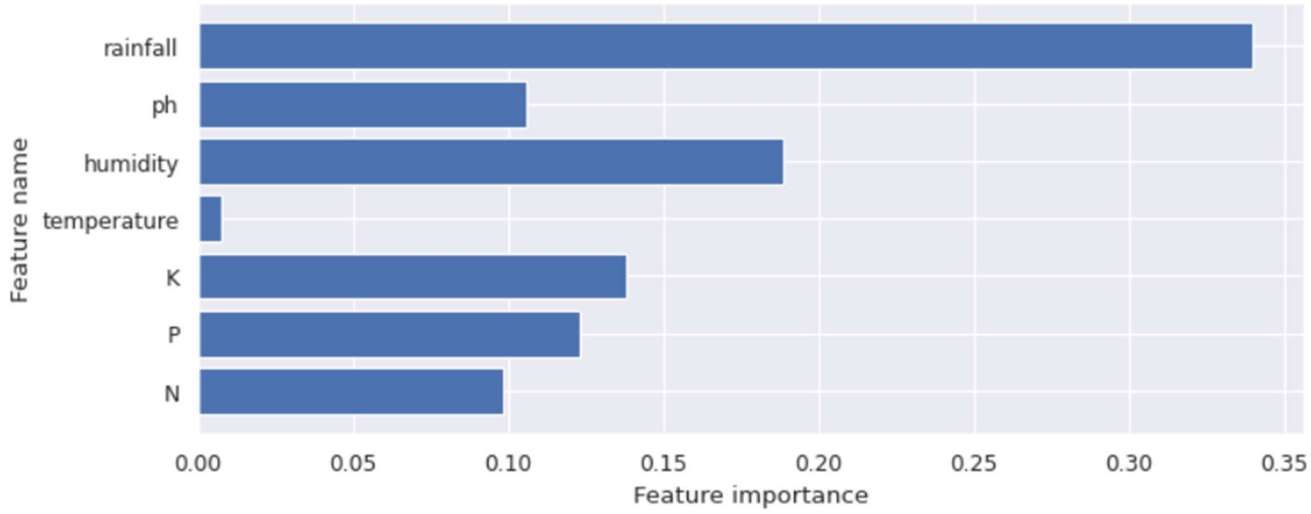


Fig. Confusion matrix

Fig. Decision Tree



7 My learnings

- Brief explanation of the Data Science AI, two career paths sometimes cross one another. Data science uses AI and its many components, including machine learning and deep learning.
- It develops and plans the whole big data environment on Hadoop and Spark. Holding experience in data visualization, data mining, and data migration is an essential requirement. Also, to have demonstrated experience with Python required to grow. It's crucial to work backward to develop the abilities required for your career, whether it's in AI or data science.
- Artificial Intelligence method works by consuming huge amounts of labeled training data. Then analyzing the data for correlations and patterns. AI uses these patterns to make predictions about business decisions. The replication of human intelligence by computer systems is defined as artificial intelligence. Speech recognition, natural language processing, and machine vision are some good examples of AI applications. Data Science is a broad process that includes pre-treatment, analysis, and visualization.
- Then it comprises prediction and generating understanding. It is the study of data to pull valuable understanding for business. It combines various disciplines to evaluate huge data.

8 Future work scope

1.Enhanced Predictive Models:

- Continuously improve predictive algorithms and models to enhance accuracy and reliability.
- Incorporate advanced machine learning techniques, such as deep learning and ensemble methods, for more robust predictions.

2.Integration of IoT and Sensor Data:

- Integrate data from IoT devices and sensors (e.g., soil moisture sensors, weather stations) to provide real-time data for more accurate predictions.
- Develop systems that can automatically collect and transmit data from these devices to predictive analytics platforms.

3.Satellite and Drone Technology:

- Explore the use of high-resolution satellite imagery and drone technology for precise monitoring of crop health and early detection of issues.
- Develop automated image recognition systems to analyze satellite and drone images efficiently.

4.Big Data Analytics:

- Leverage big data analytics to process and analyze vast amounts of agricultural data from diverse sources.
- Implement scalable infrastructure and cloud computing solutions to handle the data volume.

5.Precision Agriculture:

- Promote the adoption of precision agriculture techniques, such as variable rate technology, that use predictive analytics to optimize resource application (e.g., fertilizers, pesticides, water) based on localized conditions.