Industrial Internship Report on

"Crop and Weed Detection"

Prepared by

ELISETTY.SAI PRAVALLIKA

| 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 "Crop and Weed Detection" which aims to develop a system that only sprays pesticides on weeds and not on the crop Which will reduce the mixing problem with crops and also reduce the waste of pesticides.

This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solutions for them. It was an overall great experience to have this internship. |

Contents

1. Preface

Summary of the whole 6 weeks' work- Crop and Weed Prediction

Over the course of six weeks, our internship project on Crop Weed Prediction aimed to develop a system such that the pesticide is sprayed only on weed and not on plants . Through the fusion of data science, machine learning, and agronomic expertise, the aim is to develop an innovative solution to predict and manage weeds in a sustainable and efficient manner.
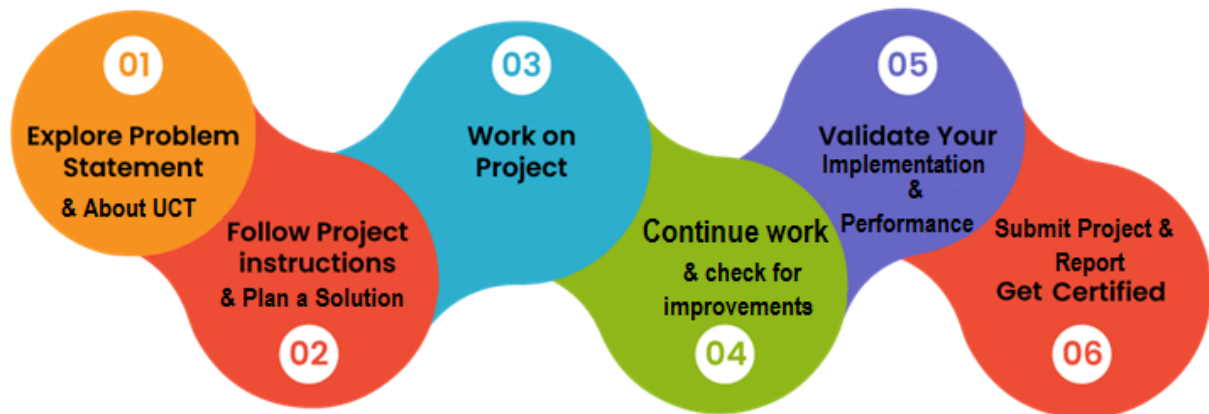
Throughout the project, extensive research and analysis is made, leveraging diverse datasets such as historical agricultural data and satellite imagery. The goal was to identify patterns and correlations between environmental factors and weed growth, that enables to build accurate predictive models. By working on such a practical and impactful project, we gained valuable hands-on experience in data analysis, machine learning, and agricultural technology. This internship provided us with a unique opportunity to apply our theoretical knowledge to real-world challenges. The problem statement for our project revolved around the detrimental impact of weeds on crop yield and food security. We recognized the need for a proactive approach to weed management, as conventional methods often rely on chemical herbicides that can have adverse effects on the environment and human health. By developing a crop weed prediction system, The aim is to provide farmers with timely and accurate information to optimize their crop management practices. This internship program, hosted by the University of Southern California (USC) and the University of Cape Town (UCT), offered a remarkable opportunity to collaborate with experts from various fields

The program was planned in such a way that ensures a comprehensive learning experience for all the participants. Exposure to various aspects of the project, including data collection, preprocessing, feature engineering, and model development. The learnings and overall experience during the internship were immensely rewarding. Valuable insights were gained regarding the potential of technology to revolutionize weed management practices.

To conclude this internship, we express our gratitude to all the mentors, supervisors, and fellow interns who contributed to our growth and success. We carry forward the knowledge and experiences gained

during this program, confident in our ability to contribute positively to the field of crop weed prediction and the broader domain of sustainable agriculture.

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.

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
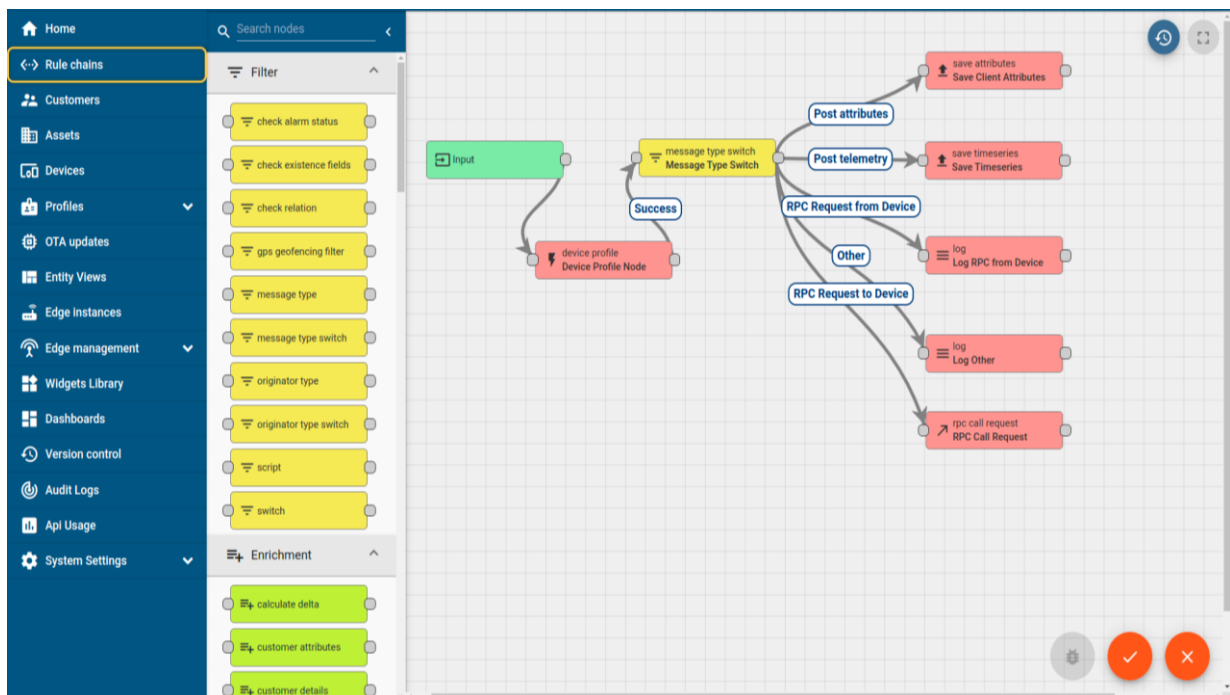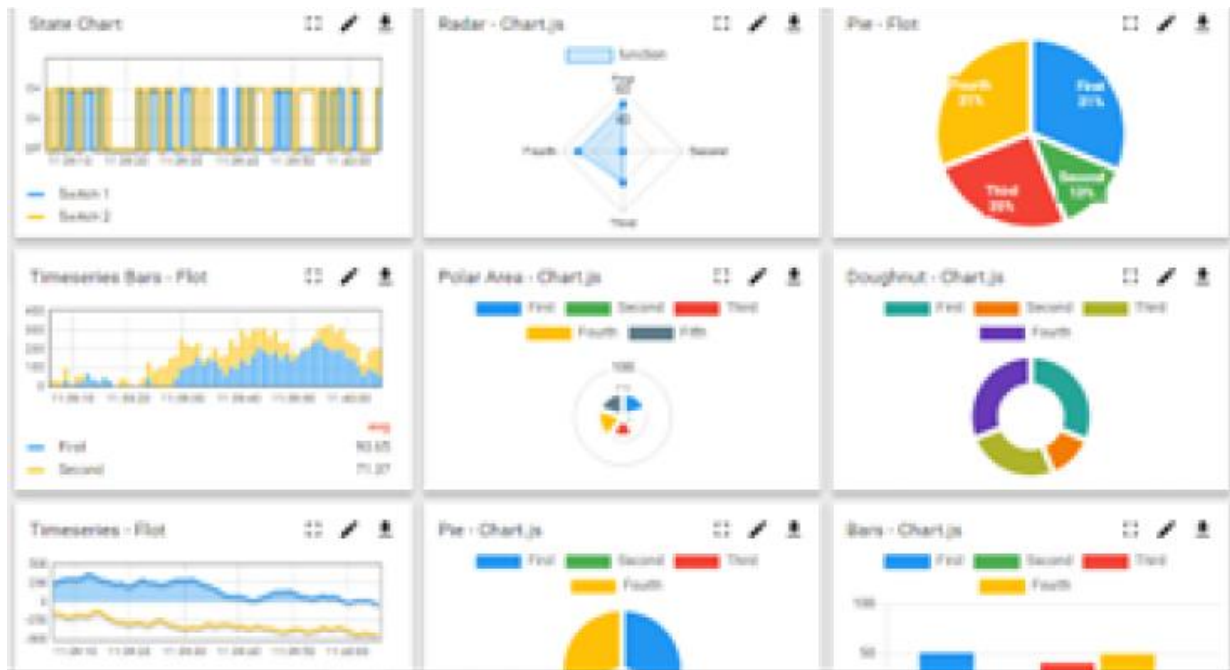
Smart Factory Platform (FACTORY WATCH)

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.

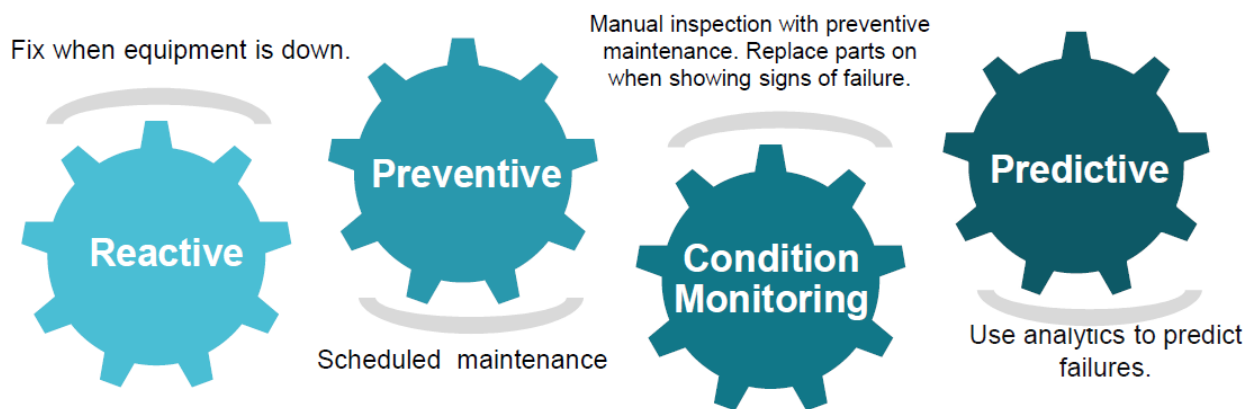| Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator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

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.

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.

Fix when equipment is down.

Manual inspection with preventive maintenance. Replace parts on when showing signs of failure.

Preventive

Reactive

Scheduled maintenance

Predictive

Condition Monitoring

Use analytics to predict failures.

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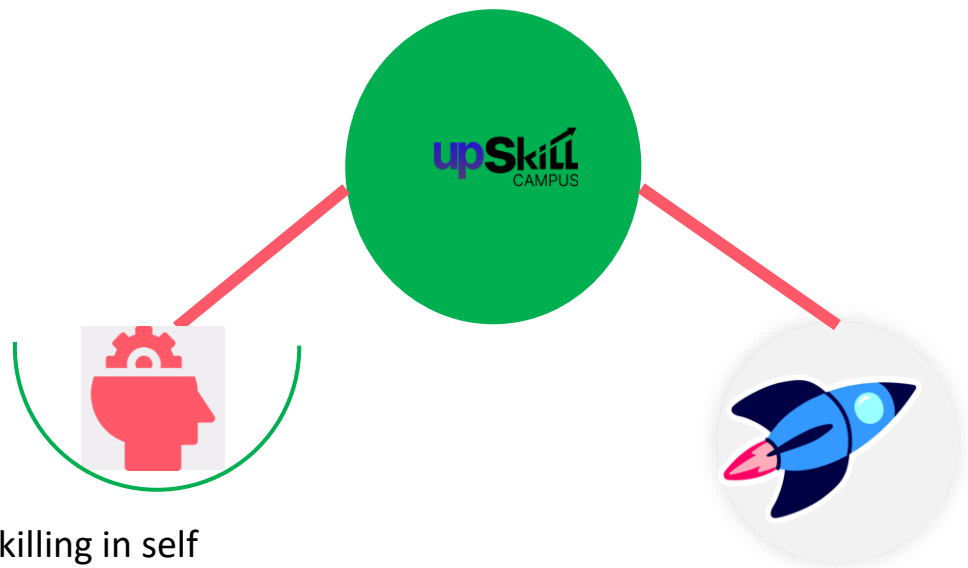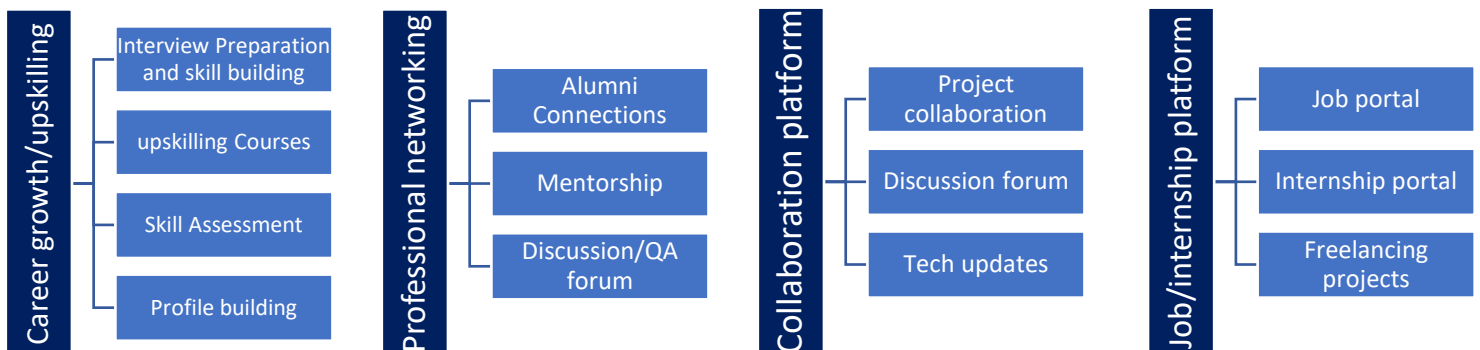
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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

https://www.upskillcampus.com/

Career growth/upskilling	Professional networking	Collaboration platform	Job/internship platform
Interview Preparation and skill building	Alumni Connections	Project collaboration	Job portal
upskilling Courses	Mentorship	Discussion forum	Internship portal
Skill Assessment	Discussion/QA forum	Tech updates	Freelancing projects
Profile building			

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. How to Train YOLO v5 on a Custom Dataset | Paperspace Blog
2. tf.keras.utils.image_dataset_from_directory | TensorFlow v2.13.0
3. Step by step VGG16 implementation in Keras for beginners | by Rohit Thakur | Towards Data Science

2.6 Glossary

Terms	Acronym
Deep Learning	DL
Vgg-16	VGG16
Convolution layer	Conv2D
Normalization	Norm
Preprocessing	Preproc

3. Problem Statement

The assigned problem statement is:

Weed is an unwanted thing in agriculture. Weed uses nutrients, water, land and many more things that might have gone to crops. Which results in less production of the required crop. The farmer often uses pesticides to remove weed which is also effective but some pesticides may stick with crop and may causes problems for humans.

We aim to develop a system that only sprays pesticides on weed and not on the crop Which will reduce the mixing problem with crops and also reduce the waste of pesticides.

4. Existing and Proposed solution

One of the existing Machine Learning algorithms used for crop weed prediction is " Random Forest" algorithm. It is an ensemble learning method that combines multiple decision trees to make predictions. In the context of crop weed prediction, the Random Forest algorithm can be trained on historical agricultural data, including information about environmental factors, crop characteristics, and weed occurrences. The algorithm learns to identify patterns and relationships between these variables and uses them to predict the presence or absence of weeds in future instances. During the training phase, the Random Forest algorithm creates a multitude of decision trees, each trained on a different subset of the data and using random feature subsets. This randomness helps to improve the algorithm's robustness and reduce the risk of overfitting. When making predictions, the Random Forest algorithm combines the outputs of all decision trees and provides a final prediction. This ensemble approach helps to mitigate the biases and uncertainties associated with individual decision trees, leading to more accurate and reliable predictions. Random Forest algorithms have the advantage of being able to handle large and complex datasets, including high-dimensional feature spaces. They are also capable of capturing nonlinear relationships between variables, making them suitable for crop weed prediction tasks where the relationships between environmental factors and weed occurrences may be complex.

Overall, the Random Forest algorithm provides a traditional machine learning approach for crop weed prediction and provides reliable predictions based on historical data and an ensemble of decision trees.

Limitations:

While the Random Forest algorithm has several advantages for crop weed prediction, it also has some limitations to consider:

1. Interpretability: Random Forest models can be challenging to interpret compared to simpler models like linear regression. The ensemble nature of the algorithm makes it difficult to understand the exact contribution of each feature to the predictions.

2. Overfitting: Although Random Forests are designed to reduce overfitting, they can still be prone to it, especially if the number of trees in the forest is too high or the model is not properly regularized. Overfitting occurs when the model becomes too complex and learns noise or idiosyncrasies in the training data, leading to poor generalization on unseen data.

3. Computational Resources: Random Forests can be computationally expensive, particularly when dealing with large datasets or high-dimensional feature spaces. The training process involves building multiple decision trees, which can require substantial memory and processing power.

4. Imbalanced Data: If the dataset used for training the Random Forest model is imbalanced, meaning the number of instances of one class (e.g., weed occurrences) is significantly smaller than the other class (e.g., non-weed occurrences), the model may be biased towards the majority class and struggle to accurately predict the minority class.

5. Feature Engineering: Random Forests generally require careful feature engineering to achieve optimal performance. Choosing relevant features and appropriately encoding categorical variables or handling missing data can significantly impact the model's accuracy.

Proposed Solution:

To address the limitations of Random Forest algorithm (RF) with a Deep Learning model, we can propose the following solution:

1. Deep Learning models can be complex, techniques such as feature importance analysis, model visualization, and attention mechanisms can provide insights into the model's decision-making process. These techniques can help improve interpretability and understanding of the model's predictions.

2. Implement regularization techniques such as dropout, L1/L2 regularization, and early stopping during the training of the Deep Learning model. These techniques help prevent overfitting by reducing the model's reliance on specific features and limiting its complexity.

3. Utilize strategies such as transfer learning and pre-trained models, such as VGG-16 or other established architectures. Transfer learning allows the model to leverage knowledge gained from large-scale datasets, accelerating the training process and improving performance, even with limited data.

4. Address the issue of imbalanced data by employing techniques like oversampling, undersampling, or generating synthetic samples using approaches like SMOTE (Synthetic Minority Over-sampling Technique). These methods help balance the distribution of the classes and mitigate bias towards the majority class.

5. Leverage the power of Deep Learning models to automatically learn relevant features from raw data. This eliminates the need for extensive manual feature engineering, reducing human bias and improving the model's ability to capture complex relationships between environmental factors and weed occurrences.

6. Deep Learning models can handle both classification and regression tasks. For continuous predictions, modify the model's output layer and loss function accordingly, allowing it to estimate continuous variables such as weed density or crop yield.

Deep Learning:

Deep learning is a subfield of machine learning that focuses on training artificial neural networks to learn and make predictions from complex data. In the context of crop weed prediction, deep learning techniques can be utilized to analyse images, sensor data, and other types of agricultural data to identify and classify weeds in crop fields.

Deep learning models, such as convolutional neural networks (CNNs), are particularly well-suited for image-based tasks like crop weed prediction. CNNs are designed to mimic the visual processing of the human brain by employing layers of interconnected artificial neurons that can learn and extract meaningful features from images.

To apply deep learning to crop weed prediction, a typical approach involves the following steps:

1. Data Collection: Gather a large dataset of labelled images that contain both crop plants and various weed species. These images should cover different weed types, growth stages, lighting conditions, and field environments.

2. Data Preprocessing: Prepare the data by resizing the images to a uniform size, normalizing pixel values, and augmenting the dataset through techniques like rotation, flipping, and zooming. These preprocessing steps help to increase the diversity and robustness of the data.

3. Model Architecture: Design a deep learning model architecture suitable for crop weed prediction. This often involves stacking multiple convolutional layers, followed by pooling layers to extract and consolidate spatial features from the images. Additional fully connected layers and an output layer are then added to make predictions.

4. Training: Train the deep learning model using the prepared dataset. This involves feeding the labelled images into the model, computing predictions, and comparing them with the ground truth labels. The model adjusts its internal parameters through a process called backpropagation, where gradients are computed and used to update the model's weights.

5. Validation and Evaluation: Assess the model's performance by validating it on a separate dataset that was not used for training. Metrics such as accuracy, precision, recall, and F1 score are used to evaluate the model's ability to correctly identify and classify weeds.

6. Deployment and Prediction: Once the model has been trained and evaluated, it can be deployed in production systems. New images of crop fields can be input into the model, and it will generate predictions indicating the presence and types of weeds in the image.

Deep learning techniques offer several advantages for crop weed prediction. They can learn intricate patterns and relationships in images, handle complex and high-dimensional data, and adapt to different field conditions. By leveraging the power of deep learning, crop weed prediction models can improve accuracy, reduce reliance on manual feature engineering, and enhance the efficiency of weed management practices in agriculture.

4.1 Code submission (Github link) :

https://github.com/Elisettysaipravallika/UpskillCampus/blob/main/Crop_weed_detection.ipynb

4.2 Report submission (Github link) :

https://github.com/Elisettysaipravallika/UpskillCampus/blob/main/CropWeedDetection_%20Elisetty_Sai
pravallika_USC_UCT.pdf

5.Performance Test

Performance testing is a critical aspect of the Crop Weed Prediction project, as it assesses the system's ability to meet specific constraints and requirements of real industries. To ensure the design accommodates these constraints, we identified the following key performance factors and implemented strategies to address them:

1. Memory Usage: We employed memory optimization techniques such as data batching and efficient storage formats to minimize the memory footprint of the deep learning models and associated data. This ensures that the system can handle large datasets and operate efficiently within limited memory constraints.

2. Computational Speed: To address the need for fast and efficient operations, we utilized optimized deep learning frameworks and libraries that leverage hardware acceleration (e.g., GPUs) to accelerate the processing of images and predictions. By leveraging parallel processing capabilities, we aimed to achieve high throughput and reduce the time required for inference.

3. Accuracy and Precision: We focused on training accurate and robust deep learning models to ensure reliable predictions of weed occurrences. We carefully curated and labeled the training data, incorporated regularization techniques to prevent overfitting, and fine-tuned the models to strike a balance between accuracy and generalization.

4. Power Consumption: While power consumption may not be a direct concern in software-based performance testing, we aimed to optimize the models and algorithms to operate efficiently on energy-constrained devices or embedded systems. By minimizing unnecessary computations and leveraging hardware acceleration, we can achieve better power efficiency.

During the performance testing phase, we conducted various experiments to measure and evaluate the system's performance against these constraints. This involved assessing memory usage, computation time, accuracy metrics, and power consumption (if applicable) under different scenarios and datasets.

The test results provided insights into the system's performance and highlighted areas for improvement. In cases where specific constraints could not be tested directly, we considered their potential impact on the design. For example, limited memory could lead to reduced model complexity or the need for data compression techniques. In such cases, we recommend optimizing the system further by exploring techniques like model compression, quantization, or pruning to reduce memory requirements.

To handle potential constraints, we recommend the following:

- Continuously monitor and profile the system's performance to identify any bottlenecks or areas requiring optimization.

- Utilize resource monitoring tools to track memory usage, computation time, and power consumption in real-world deployments.

- Employ techniques like distributed computing or model parallelism to scale up the system's performance and handle larger datasets or higher prediction loads.

- Collaborate with hardware experts to explore hardware-specific optimizations or dedicated hardware solutions to further enhance performance.

By proactively addressing and monitoring these constraints, we can ensure that the Crop Weed Prediction system delivers accurate predictions, operates efficiently, and meets the performance requirements of real industries.

6.My learnings

Participating in the Crop Weed Prediction project has been a transformative experience, providing valuable insights and growth opportunities that will significantly impact my career development. Here is a summary of my overall learning and how it will contribute to my career growth:

1. Technical Skills Enhancement: Throughout the project, I acquired and honed various technical skills, including data preprocessing, deep learning model development, image analysis, and performance optimization. These skills have not only equipped me with in-depth knowledge in crop weed prediction but have also strengthened my proficiency in data science and machine learning, positioning me as a competent professional in the field.

2. Domain Expertise in Agriculture: Working on the Crop Weed Prediction project allowed me to delve into the domain of agriculture and gain a deep understanding of the challenges and complexities involved in crop management and weed control. This expertise will be invaluable as I pursue a career in the agricultural industry, enabling me to provide data-driven solutions and contribute to sustainable and efficient agricultural practices.

3. Multidisciplinary Collaboration: The project's multidisciplinary nature provided me with the opportunity to collaborate with experts from various fields, including data scientists, agronomists, and software engineers. This experience enhanced my ability to work effectively in diverse teams, fostered an appreciation for different perspectives, and improved my communication and teamwork skills, which are essential for success in any professional setting.

4. Real-World Application: Developing a practical solution for crop weed prediction emphasized the importance of bridging the gap between theory and real-world application. I gained hands-on experience in addressing complex problems, understanding business requirements, and translating them into actionable solutions. This practical knowledge will serve as a strong foundation for future projects and strengthen my ability to deliver tangible results in professional settings.

5. Continuous Learning and Adaptability: The project exposed me to the dynamic and rapidly evolving field of deep learning and agricultural technology. The constant exploration of new techniques, tools, and methodologies emphasized the importance of continuous learning and adaptability in staying at the forefront of the industry. It instilled in me a passion for lifelong learning and a mindset of embracing change and innovation.

7.Future work scope

We were unable to thoroughly investigate various ideas and approaches during the Crop Weed Prediction research due to time constraints. These concepts, however, have potential and can be followed in the future to improve the system. Here are a few examples:

Multi-Sensor Fusion: By combining data from many sensors, such as satellite imaging, drones, and IoT devices, a more comprehensive and accurate picture of the agricultural landscape can be obtained. We can improve weed prediction models and obtain better insights into the interplay of environmental factors and weed occurrences by merging data from various sources.

Long-Term Weed Growth Forecasting: Moving beyond short-term weed occurrences to long-term weed growth forecasting would allow farmers to plan and implement proactive weed management measures. We can develop models that predict weed growth trends and enhance long-term decision-making by taking historical data, weather patterns, and crop rotations into account.