Industrial Internship Report on

Crop Yield Production Prediction

Prepared by

Sanjivani Hinesh Rajput

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 focused on crop yield production prediction. The main objective of the project was to develop a predictive model that could estimate crop yields based on various factors such as weather condition like rainfall , and other relevant parameters. The ultimate goal was to provide farmers and agricultural stakeholders with accurate predictions of crop yields, enabling them to make informed decisions regarding planting, resource allocation, and crop management strategies. 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

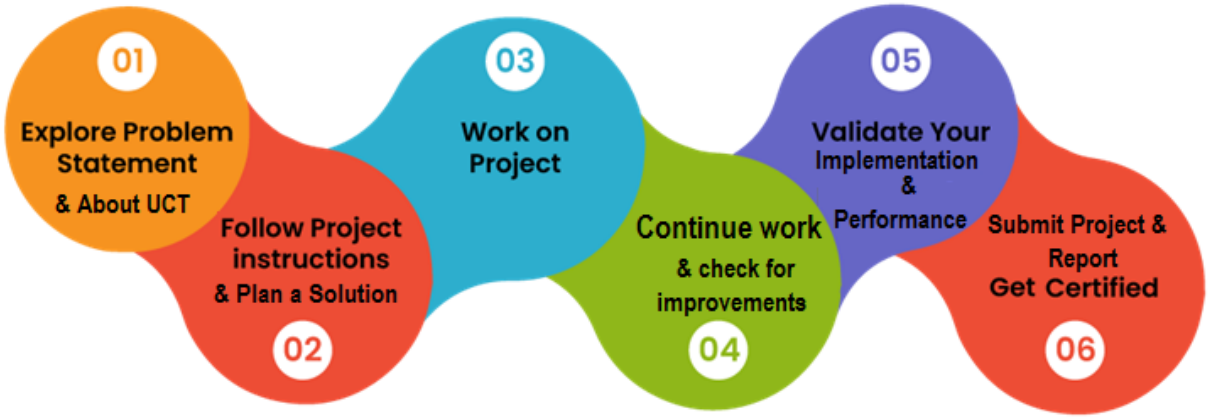
Summary of the whole 6 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

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.

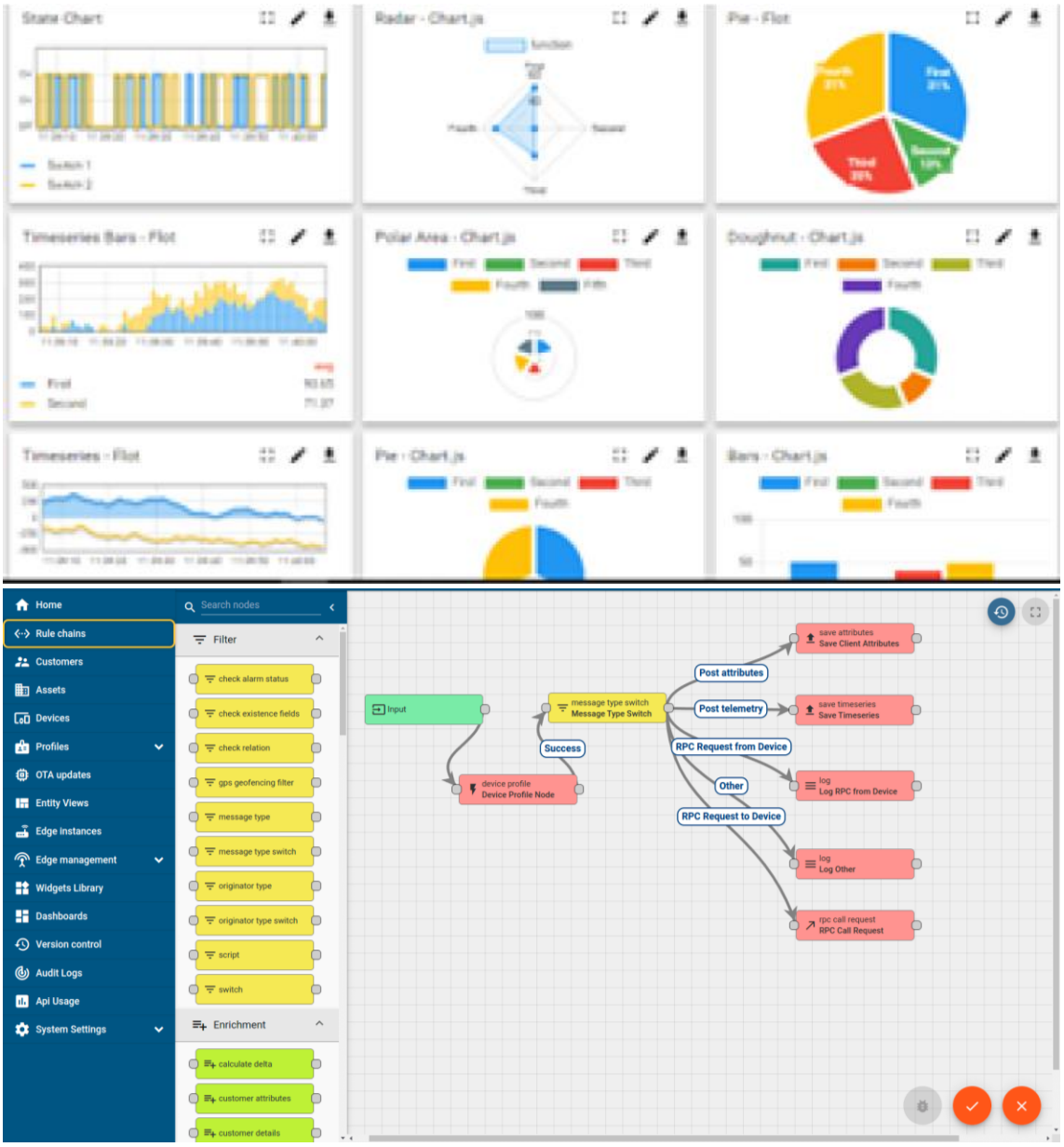
i. UCT IoT Platform (UCT Insight)

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.

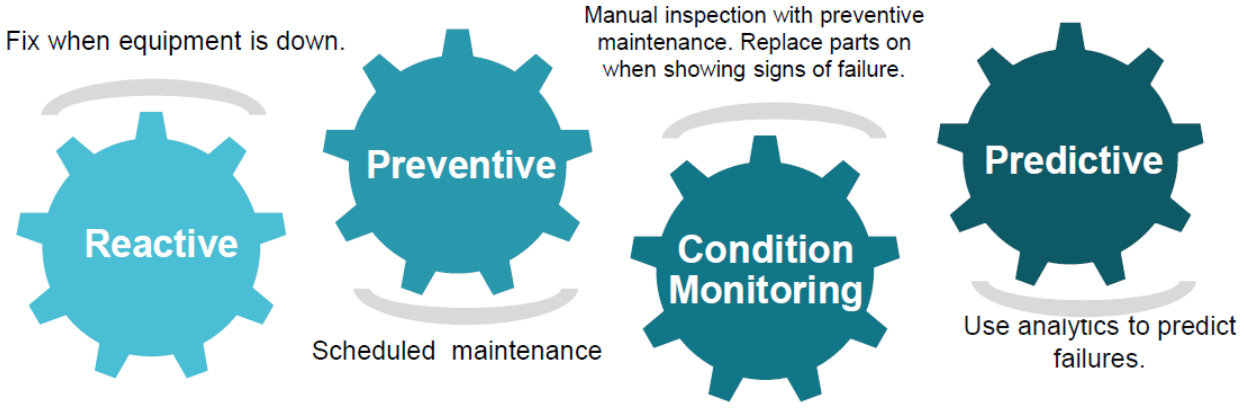
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iv. Predictive Maintenance

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



2.2 About upskill Campus (USC)

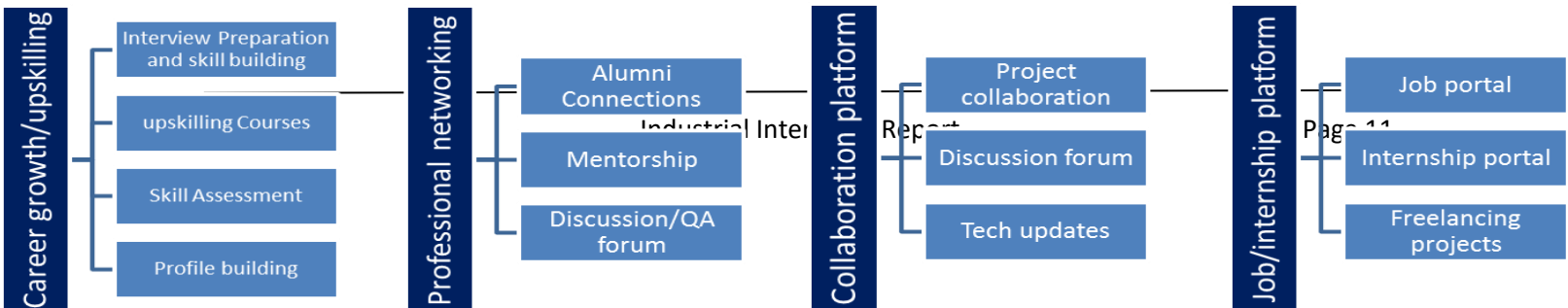
upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

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

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

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

https://www.upskillcampus.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] datafile.csv

[2] https://data.gov.in/

[3] Kaggle

2.6 Glossary

Terms	Acronym
Crop	
Yield	
Rainfall	
Decision tree	

3 Problem Statement

In the assigned problem statement

The problem statement for this project revolves around developing an efficient and accurate predictive model for crop yield production. The objective is to leverage historical data, weather conditions, and other relevant factors to forecast crop yields in a given agricultural region. By addressing this challenge, we aim to empower farmers and stakeholders with valuable insights to make informed decisions on crop planning, resource allocation, and agricultural management strategies. The successful implementation of the predictive model will have a significant impact on enhancing farming practices, optimizing productivity, and contributing to the overall sustainability of the agricultural sector.

4 Existing and Proposed solution

EXISTING SOLUTION

There are several existing solutions and methods for predicting crop production. Here are some of the common existing approaches:

1. Expert Knowledge: In many agricultural regions, experienced farmers and agricultural experts use their knowledge and experience to estimate crop production based on factors like crop variety, weather patterns, soil quality, and local practices. While this approach can be effective, it is subjective and may not account for complex interactions between various variables.

2. Statistical Models: Some existing solutions use statistical models like linear regression or time series analysis to analyze historical crop yield data along with weather and soil data. These models attempt to identify patterns and relationships between variables to make predictions about future crop yields. However, they might not capture all the nuances of the agricultural ecosystem.

3. Remote Sensing and Satellite Imagery: Modern technologies like remote sensing and satellite imagery are increasingly used to monitor and assess crop health, growth, and yield potential. By analyzing satellite data, researchers and agricultural experts can identify crop stress, assess vegetation indices, and make inferences about crop productivity.

4. Crop Simulation Models: Crop simulation models, such as the widely used DSSAT (Decision Support System for Agrotechnology Transfer), aim to simulate crop growth and development based on environmental factors. These models simulate the effects of various management practices and environmental conditions on crop yields.

5. Machine Learning and AI: With advancements in machine learning and artificial intelligence, data-driven models have become more prevalent in crop yield prediction. Machine learning algorithms, such as decision trees, random forests, and

neural networks, can learn complex patterns from large datasets and improve prediction accuracy.

PROPOSED SOLUTION

The data contains various fields that can be used to link other datasets. After doing that we clean the dataset and remove the missing variables and normalized the data. Then after we split the dataset into training and testing. The model is trained using the training dataset and checked on the testing dataset. Since this contains more than two output variables linear regression fails. Random Forest provides better results but it is slow. Thus after using the Decision Tree, we can get the desired results.

4.1 Code submission (Github link)

spiderrex/Crop-yield-production: This predicts the crop yield production through out India using Decision Tree (github.com)

4.2 Report submission (Github link) :.

5 spiderrex/Crop-yield-production: This predicts the crop yield production through out India using Decision Tree (github.com)

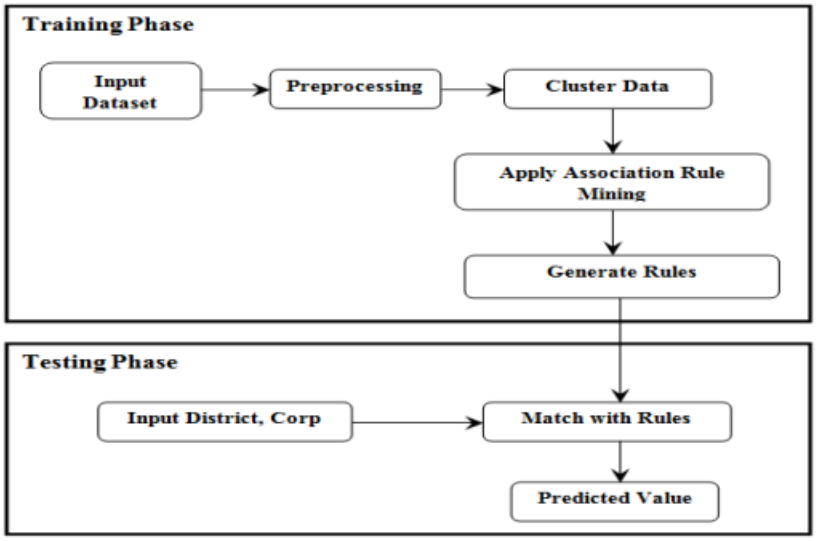


Figure 1. Proposed Work Flow

In this project, our system harnesses the power of Machine Learning techniques to predict the rate of crop yield. Python, being widely acclaimed for its versatility in the field of Machine Learning, serves as our programming language of choice for implementing novel ideas. By leveraging collected data sets, we aim to generate precise crop yield predictions using well-known Machine Learning algorithms such as Random Forest Regressor and Decision Tree Regressor. The accuracy of our predictions depends heavily on the richness and relevance of the information contained within the collected data sets. The more comprehensive and detailed the parameters in our data sets, the better our predictions are likely to be. Through this approach, we aspire to provide farmers and agricultural stakeholders with valuable insights to optimize crop planning, resource allocation, and overall agricultural productivity.

METHODOLOGY

Data Pre-Processing

Data Preprocessing is a method that is used to convert the raw data into a clean data set. The data are gathered from different sources, it is collected in raw format which is not feasible for the analysis. By applying different techniques like replacing missing values and null values, we can transform data into an understandable format. We use one hot encoding on District_Name to convert it into numerical values. Then we add 'yield' field as target output.

The final step on data preprocessing is the splitting of training and testing data. The data usually tend to be split unequally because training the model usually requires as much data- points as possible. The training dataset is the initial dataset used to train ML algorithms to learn and produce right predictions Factors affecting Crop Yield and Production(Here it is splitted in 75-25)

```
# Dropping missing values
crop_data = crop_data.dropna()
crop_data
```

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

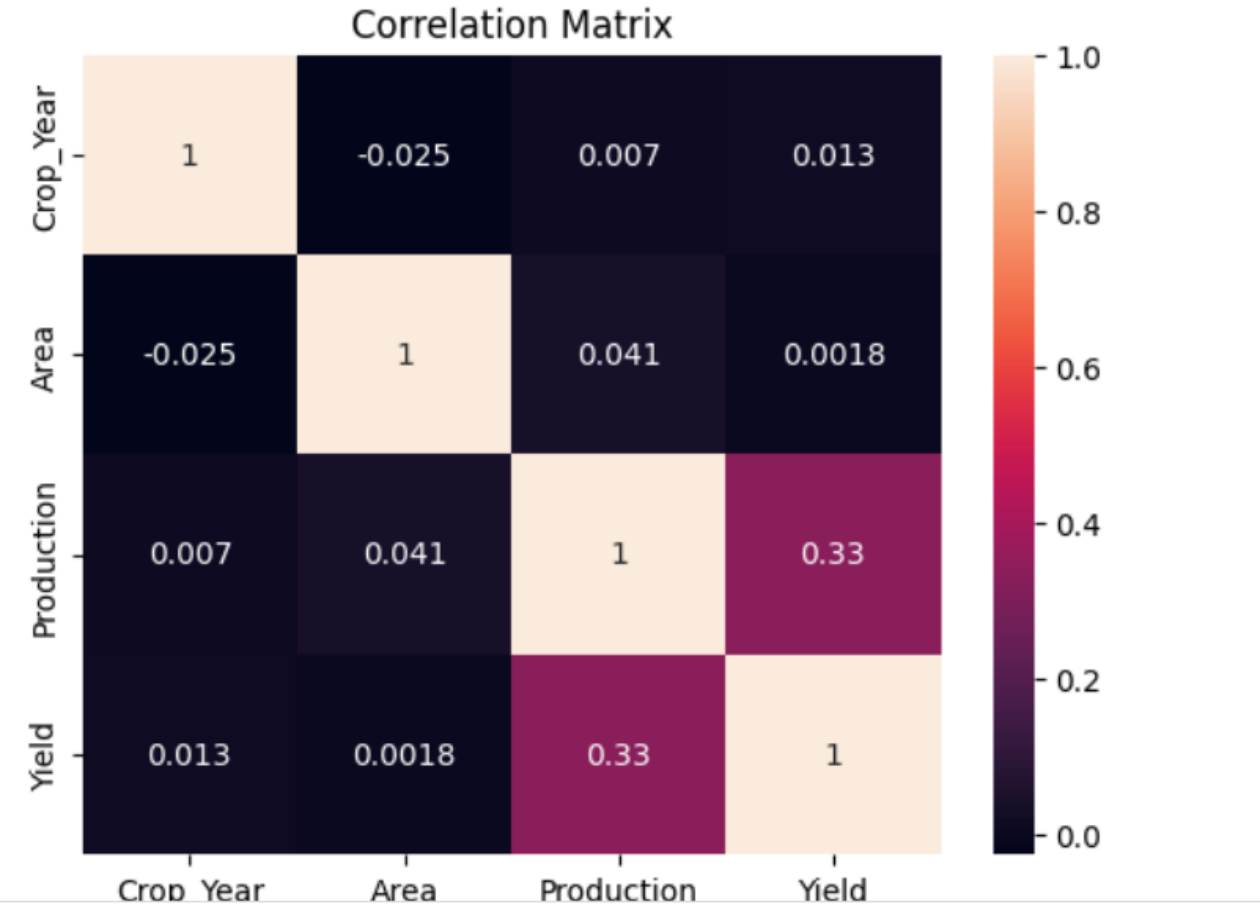
242361 rows × 7 columns

```
# Adding a new column Yield which indicates Production per unit Area.

crop_data['Yield'] = (crop_data['Production'] / crop_data['Area'])
crop_data.head(10)
```

```
sns.heatmap(data.corr(), annot =True)
plt.title('Correlation Matrix')
```

```
<ipython-input-13-3c0a03b5e131>:1: FutureWarning: The default value of nume
  sns.heatmap(data.corr(), annot =True)
Text(0.5, 1.0, 'Correlation Matrix')
```



Comparison and Selection of Machine Learning Algorithm

Before deciding on an algorithm to use, first we need to evaluate and compare, then choose the best one that fits this specific dataset. Machine Learning is the best technique which gives a better practical solution to crop yield problem. There are a lot of machine learning algorithms used for predicting the crop yield. In this paper we include the following machine learning algorithms for selection and accuracy comparison :

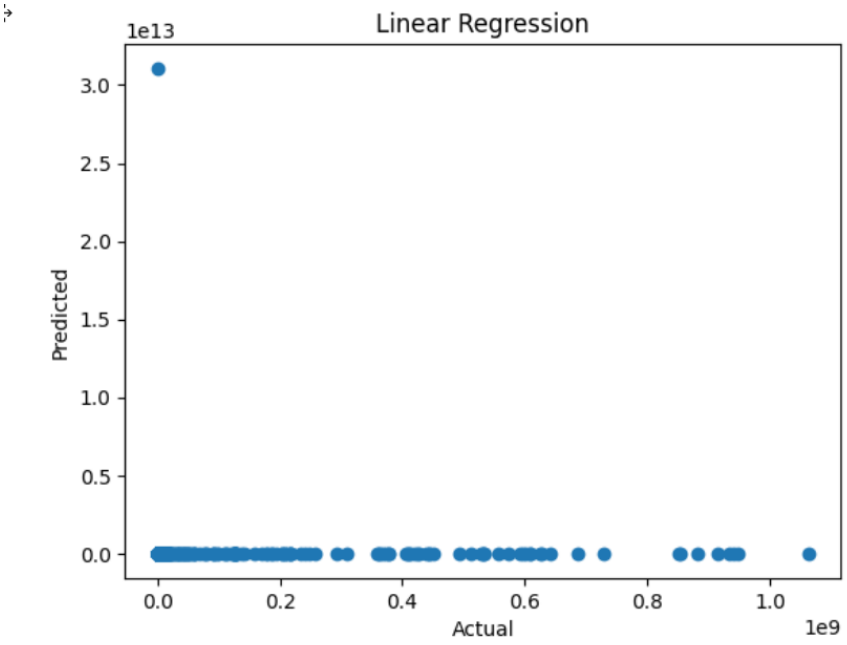
1 .Linear Regression:-When linear regression applied the results weren't that accurate.

```
[19] model.score(x_test,y_test)

     -55393024.150060125
```

```
[20] from sklearn.metrics import r2_score
     r = r2_score(y_test,lr_predict)
     print("R2 score : ",r)

     R2 score :  -55393024.150060125
```



2. Random Forest:- Random Forest has the ability to analyze crop growth related to the current climatic conditions and biophysical change. Random Forest provided better results but it was quite slow.

```
model.score(x_test,y_test)
```

```
0.9646900995456444
```

```
# Calculating R2 score

from sklearn.metrics import r2_score
r1 = r2_score(y_test,rf_predict)
print("R2 score : ",r1)
```

```
R2 score :  0.9646900995456444
```

3. Decision Tree: Decision Tree provided better results and was faster than random forest.

```
regressor.score(x_test,y_test)
```

```
0.9564181540365296
```

```
# Calculating R2 score :

from sklearn.metrics import r2_score
r21 = r2_score(y_test,decisiontree_predict)
print("R2 score : ",r21)
```

```
R2 score :  0.9564181540365296
```

```
# Calculating Adj. R2 score:

Adjr2_2 = 1 - (1-r21)*(len(y_test)-1)/(len(y_test)-x_test.shape[1]-1)
print("Adj. R-Squared : {}".format(Adjr2_2))
```

5 Performance Test

```
# Make predictions on the test set
y_pred = final_dt_regressor.predict(x_test)
```

```
# Calculate the RMSE and R-squared on the test set
test_rmse = np.sqrt(mean_squared_error(y_test, y_pred))
test_r2 = r2_score(y_test, y_pred)
print("Test RMSE:", test_rmse)
print("Test R-squared:", test_r2)
```

```
Test RMSE: 3624380.5323943435
Test R-squared: 0.8943757344909651
```

6 My learnings

In this project, I leverage the capabilities of Machine Learning techniques to tackle the challenge of crop yield prediction. Python, being a widely accepted language in the realm of Machine Learning, serves as our primary tool for implementing innovative ideas. By utilizing carefully curated data sets, we aim to generate accurate crop yield predictions through the application of well-established Machine Learning algorithms like Random Forest Regressor and Decision Tree Regressor. The success of our predictions heavily relies on the quality and comprehensiveness of the information present in the collected data sets. The more insightful and detailed the parameters captured, the more refined our predictions will be. Our ultimate goal is to provide farmers and agricultural stakeholders with reliable insights to optimize agricultural practices, make informed decisions, and enhance overall crop productivity. With this project, we endeavor to contribute to the advancement of agricultural technology and the sustainable growth of the farming community.

7 Future work scope

The proposed system aims to address the pressing issue of increasing farmer suicides by empowering farmers to achieve financial stability and success. Through the Crop Recommender system, farmers can predict the yield of specific crops and make informed decisions about what crops to cultivate. Additionally, the system offers valuable insights on the optimal timing for fertilizer application, ensuring efficient resource utilization.

To achieve this, we collected and meticulously studied relevant datasets, which were then utilized for training our advanced machine learning tools. By combining the power of data-driven algorithms with real-time location tracking, the system dynamically fetches crucial information from the backend based on the user's location. This user-friendly approach reduces the burden on farmers as they only need to provide limited information such as soil type and area. By doing so, we aim to create a practical and effective solution that helps farmers maximize their agricultural productivity, make sustainable choices, and ultimately strengthen their financial standing. Through the Crop Recommender system, we hope to make a positive impact on the lives of farmers and contribute to the reduction of farmer distress and suicides.