Industrial Internship Report on

” Prediction of Agriculture Crop Production in India”

Prepared by,

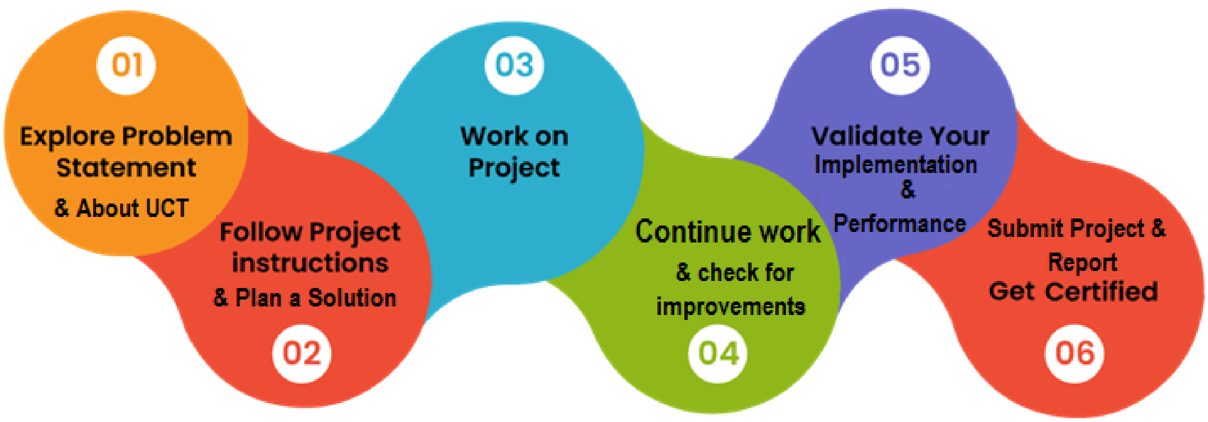
NO.	Name
1.	Rajendra M. Chaudhari
2.	Denil J. Patel
3.	Pranali Patil
4.	Chirag Gangani

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 where farming techniques are mostly obsolete bringing in technological reformations is the need of an hour. India has a diverse geographic topology. Innumerable crops hail oddly throughout the Indian land, which makes it tough to estimate the potential of any yield. For making framing more efficient and profitable we have try to create model which can predict the which type of crop is more beneficial to produce at the time. This model will take some past data and with some analysis and observations by using concept of Machine Learning and Data Science it will best crop for production for farmer 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 Content

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.
- Your message to your juniors and peers.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

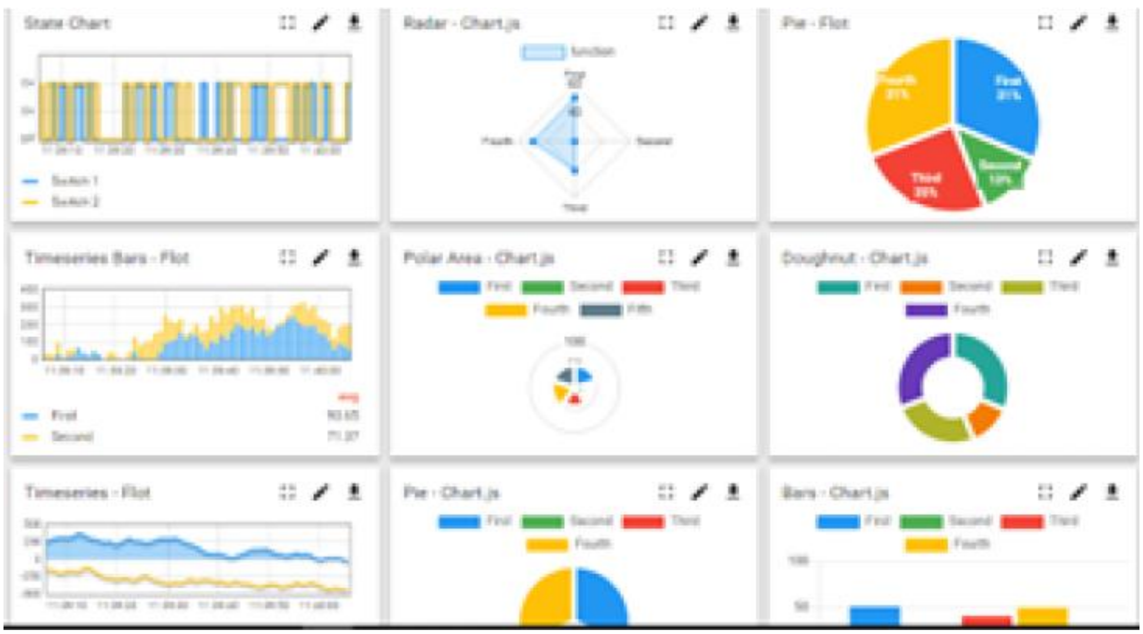
A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

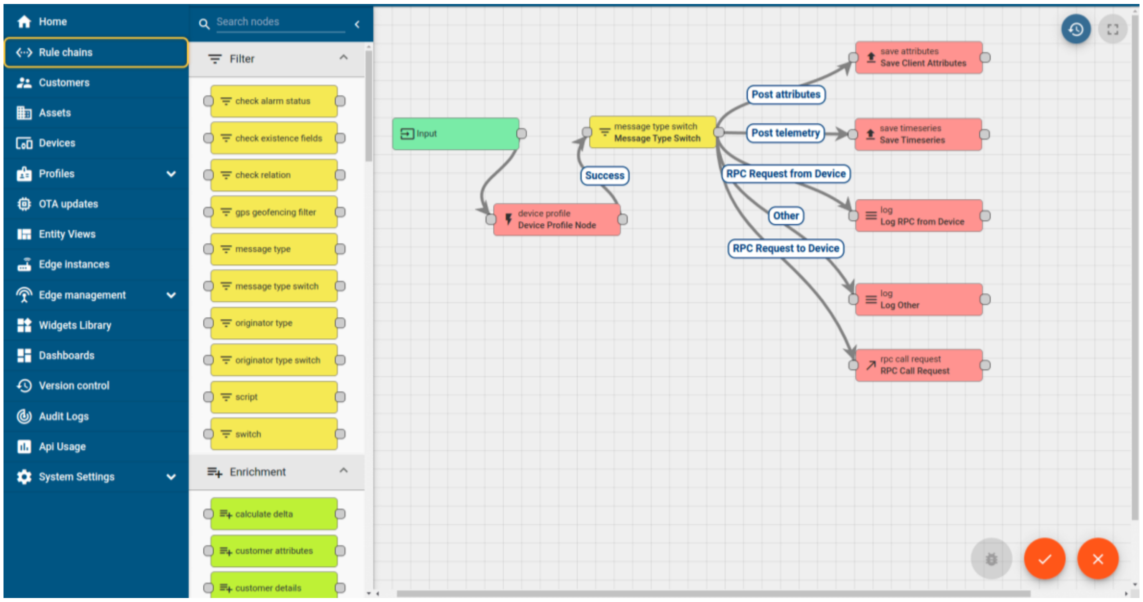
For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

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.

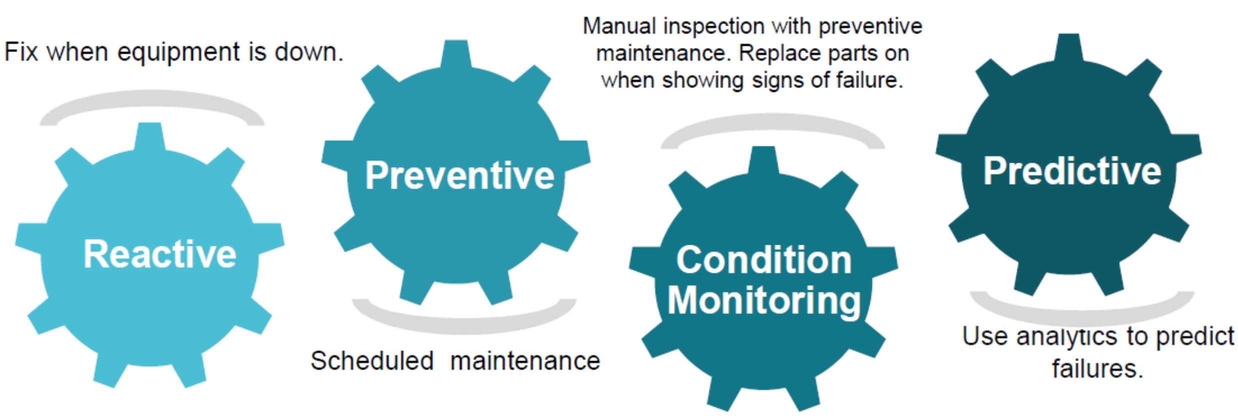
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iv. Predictive Maintenance

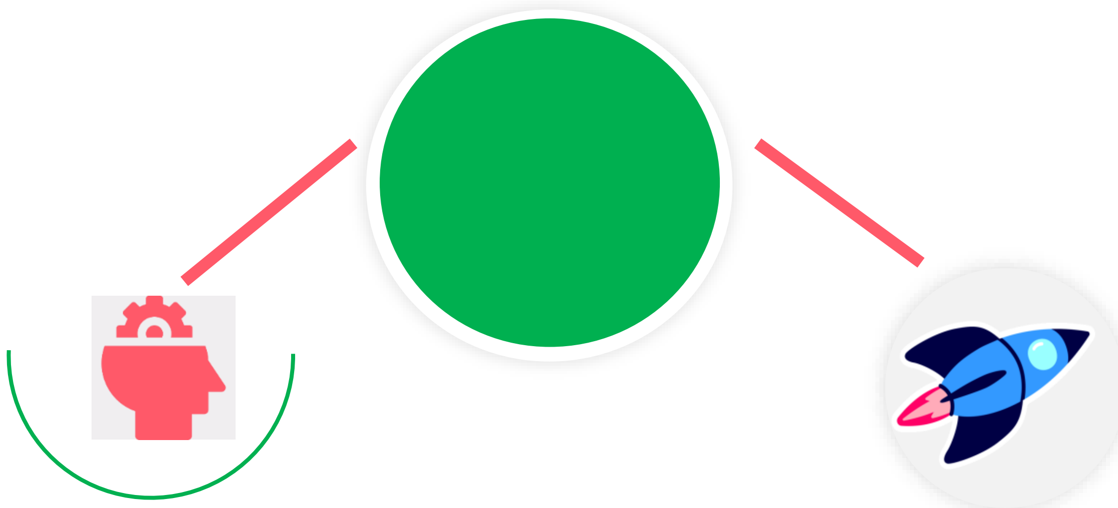
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

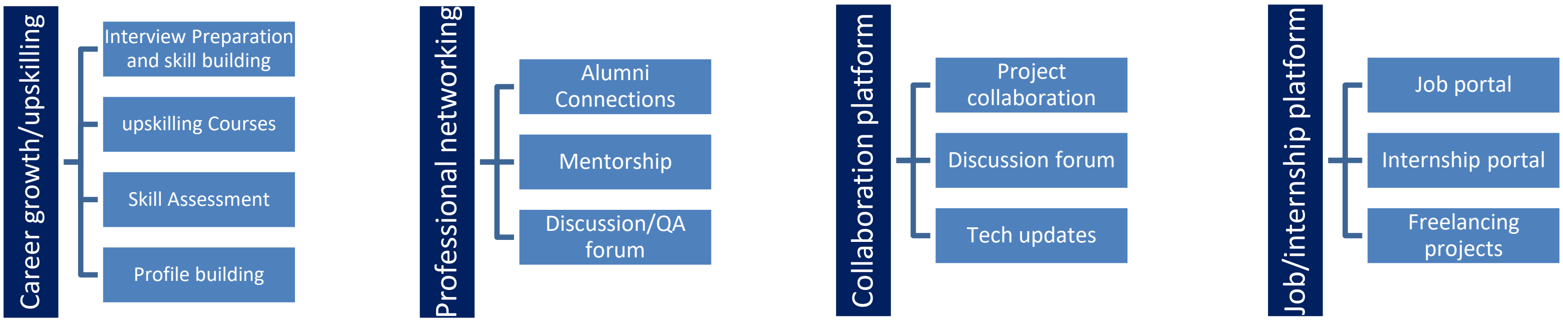
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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/



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. www.sciencedirect.com
2. www.academia.edu
3. www.researchgate.net
4. www.tandfonline.com

3 Problem Statement

3.1 Create model for prediction of Agriculture Crop Production In India :

- **Description**

Agriculture is the backbone of any country. For a country like India, where farming techniques are mostly obsolete bringing in technological reformations is the need of an hour. India has a diverse geographic topology. Innumerable crops hail oddly throughout the Indian land, which makes it tough to estimate the potential of any yield. For making framing more efficient and profitable we have try to create model which can predict the which type of crop is more beneficial to produce at the time. This model will take some past data and with some analysis and observations by using concept of Machine Learning and Data Science it will best crop for production for farmer.

3.2 What we learn by studying this problem statement

- Learn basics of Data Science and Machine Learning through the lectures provide in week 1 internship program by UCT.
- Learned types of machine learning algorithms to analyze the data and identify patterns.
- we learned from developed foreign countries that they use cameras, to measuring meters for mineral and Ph measuring. These Device use sensors embedded in the farming field to detect the content of field in terms of minerals and other required components.
- By using data collected by devices microcontroller process this data and provide output which you to predict the crop for farming

4 Existing and Proposed solution

- our project concentrating on Prediction of Agriculture Crop Production & identification, major advances were made in terms of data collecting and initial model formation.
- The team will focus on addressing challenges specific to Agriculture Crop Production & identification, which often involves distinguishing between closely resembling plant species.
- in our Prediction of Agriculture Crop Production project. The successful data collection, preprocessing, and initial model development set a solid foundation for the upcoming weeks of work. With a clear roadmap and continued collaboration, we are confident in our ability to deliver a robust and accurate detection solution to benefit the agriculture industry

4.1 Code submission

https://github.com/pranali45/Upskillcampus.git

4.2 Report submission

"Prediction of Agriculture Crop Production in India_pranali_USC_UCT.pdf

5 Proposed Design/ Model

Our proposed system is a mobile application which predicts name of the crop as well as calculate its corresponding yield. Name of the crop is determined by several features like temperature, humidity, wind-speed, rainfall etc. and yield is determined by the area and production. In this paper, Random Forest classifier is used for prediction. It will attain the crop prediction with best accurate values.

5.1 Data Preprocessing:

- Load the dataset using pandas.'
- Handle missing values, if any, in the dataset.
- Convert the "Season" column to the appropriate format (number of days)
- Convert categorical variables (e.g., "Variety," "state," "Recommended Zone") into numerical representations using techniques like one-hot encoding

5.2 Feature Selection and Engineering

- Decide which features are relevant for predicting crop production. Some potentially relevant features might be "Variety," "state," "Season," "Recommended Zone," and "Cost."
- Engineer new features if possible, such as extracting the year from the "production" column.

5.3 Data Splitting

- Split the dataset into training and testing sets. A common split ratio is 80% for training and 20% for testing

```python
crop.head()
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

```python
fert.head()
```

	Unnamed: 0	Crop	N	P	K	pH
0	0	Rice	80	40	40	5.5
1	1	Jowar(Sorghum)	80	40	40	5.5
2	2	Barley(JAV)	70	40	45	5.5
3	3	Maize	80	40	20	5.5
4	4	Ragi(naachnnii)	50	40	20	5.5

```python
crop.head()
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

```python
crop.tail()
```

	temperature	humidity	ph	rainfall	label
3095	25.287846	89.636679	6.765095	58.286977	watermelon
3096	26.638386	84.695469	6.189214	48.324286	watermelon
3097	25.331045	84.305338	6.904242	41.532187	watermelon
3098	26.897502	83.892415	6.463271	43.971937	watermelon
3099	26.986037	89.413849	6.260839	58.548767	watermelon

```python
# using extract labesl on crop to get all the data related to those labels
new_crop = pd.DataFrame(columns = crop.columns)
new_fert = pd.DataFrame(columns = fert.columns)
```

```python
for label in extract_labels:
    new_crop = new_crop.append(crop[crop['label'] == label])
```

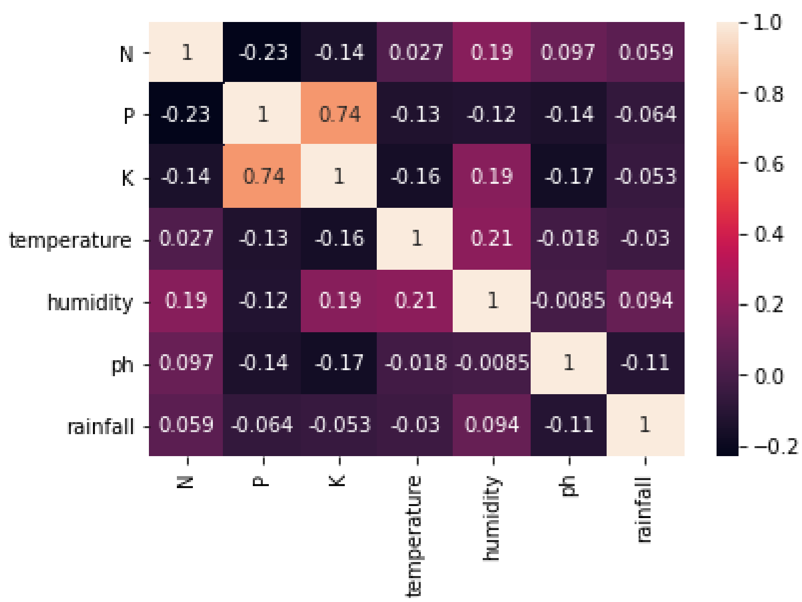
```python
for label in extract_labels:
    new_fert = new_fert.append(fert[fert['Crop'] == label].iloc[0])
```

```
new_crop
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...
895	26.774637	66.413269	6.780064	177.774507	coffee
896	27.417112	56.636362	6.086922	127.924610	coffee
897	24.131797	67.225123	6.362608	173.322839	coffee
898	26.272418	52.127394	6.758793	127.175293	coffee
899	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 5 columns

	Crop	N	P	K	pH
0	rice	80	40	40	5.5
3	maize	80	40	20	5.5
5	chickpea	40	60	80	5.5
12	kidneybeans	20	60	20	5.5
13	pigeonpeas	20	60	20	5.5
14	mothbeans	20	40	20	5.5
15	mungbean	20	40	20	5.5
18	blackgram	40	60	20	5.0
24	lentil	20	60	20	5.5
60	pomegranate	20	10	40	5.5
61	banana	100	75	50	6.5
62	mango	20	20	30	5.0
63	grapes	20	125	200	4.0
66	watermelon	100	10	50	5.5
67	muskmelon	100	10	50	5.5
69	apple	20	125	200	6.5
74	orange	20	10	10	4.0
75	papaya	50	50	50	6.0
88	coconut	20	10	30	5.0
93	cotton	120	40	20	5.5
94	jute	80	40	40	5.5
95	coffee	100	20	30	5.5



5.4 Model Selection

- Choose a regression algorithm. Let's use a Random Forest Regression model as an example.
- Import the necessary libraries (e.g., scikit-learn) and initialize the model.

5.4.1 Decision Tree

DecisionTrees's Accuracy is: 90.0

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	13
banana	1.00	1.00	1.00	17
blackgram	0.59	1.00	0.74	16
chickpea	1.00	1.00	1.00	21
coconut	0.91	1.00	0.95	21
coffee	1.00	1.00	1.00	22
cotton	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	18
jute	0.74	0.93	0.83	28
kidneybeans	0.00	0.00	0.00	14
lentil	0.68	1.00	0.81	23
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	26
mothbeans	0.00	0.00	0.00	19
mungbean	1.00	1.00	1.00	24
muskmelon	1.00	1.00	1.00	23
orange	1.00	1.00	1.00	29
papaya	1.00	0.84	0.91	19
pigeonpeas	0.62	1.00	0.77	18
pomegranate	1.00	1.00	1.00	17

5.4.2 Naive Bayes

```
Naive Bayes's Accuracy is:  0.990909090909091
              precision    recall  f1-score   support


       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       1.00      1.00      1.00        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.88      1.00      0.93        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       1.00      1.00      1.00        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
   mothbeans       1.00      1.00      1.00        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      1.00      1.00        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       1.00      0.75      0.86        16
  watermelon       1.00      1.00      1.00        15
...
    accuracy                           0.99       440
   macro avg       0.99      0.99      0.99       440
weighted avg       0.99      0.99      0.99       440
```

5.4.3 Support Vector Machine (SVM)

```
SVM's Accuracy is:  0.9795454545454545
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       1.00      1.00      1.00        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      0.95      0.98        22
      cotton       0.95      1.00      0.98        20
      grapes       1.00      1.00      1.00        18
        jute       0.83      0.89      0.86        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       1.00      1.00      1.00        23
       maize       1.00      0.95      0.98        21
       mango       1.00      1.00      1.00        26
   mothbeans       1.00      1.00      1.00        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      1.00      1.00        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       0.80      0.75      0.77        16
  watermelon       1.00      1.00      1.00        15
...
    accuracy                           0.98       440
   macro avg       0.98      0.98      0.98       440
weighted avg       0.98      0.98      0.98       440
```

5.4.4 Logistic Regression

```
Logistic Regression's Accuracy is:  0.9522727272727273
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       0.86      0.75      0.80        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       0.86      0.90      0.88        20
      grapes       1.00      1.00      1.00        18
        jute       0.84      0.93      0.88        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       0.88      1.00      0.94        23
       maize       0.90      0.86      0.88        21
       mango       0.96      1.00      0.98        26
   mothbeans       0.84      0.84      0.84        19
    mungbean       1.00      0.96      0.98        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      0.95      0.97        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       0.85      0.69      0.76        16
  watermelon       1.00      1.00      1.00        15
...
    accuracy                           0.95       440
   macro avg       0.95      0.95      0.95       440
weighted avg       0.95      0.95      0.95       440
```

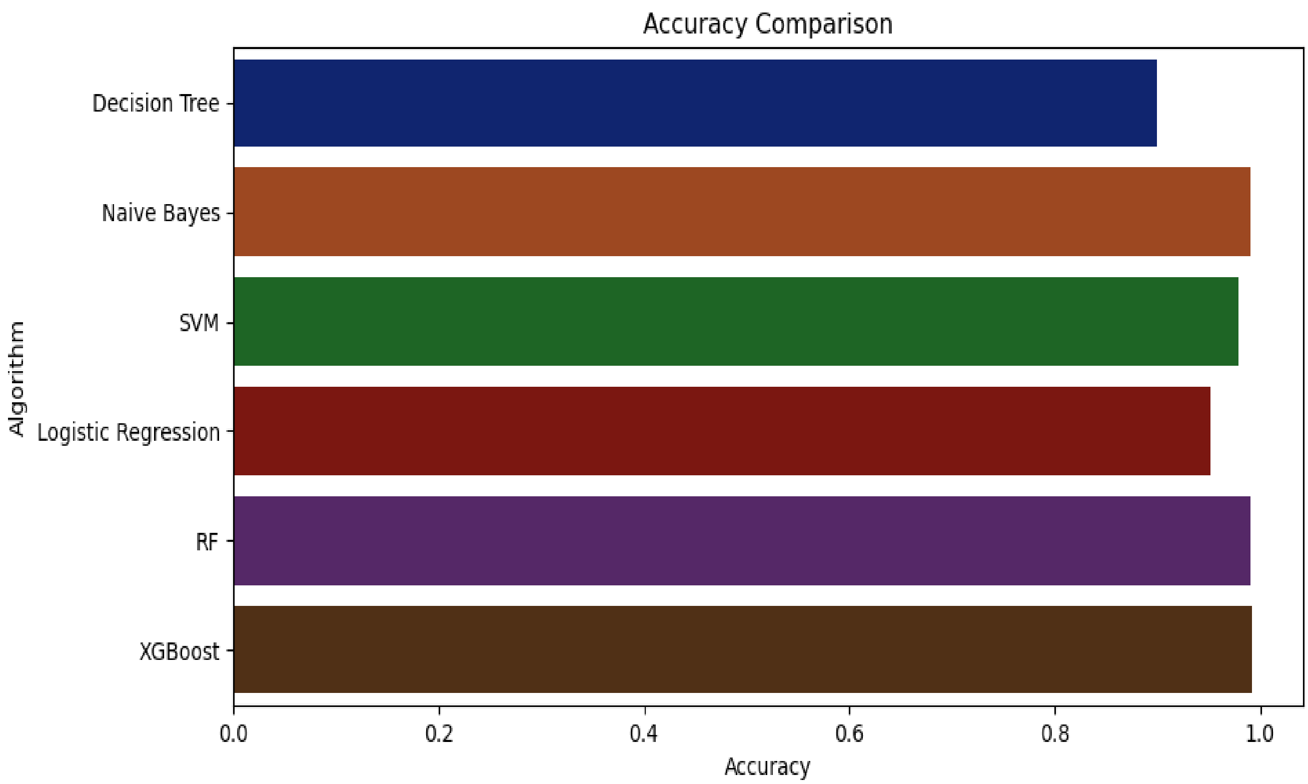
5.4.5 Random Forest

```
RF's Accuracy is:  0.990909090909091
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       0.94      1.00      0.97        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.90      1.00      0.95        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       1.00      1.00      1.00        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
   mothbeans       1.00      0.95      0.97        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      1.00      1.00        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       1.00      0.81      0.90        16
  watermelon       1.00      1.00      1.00        15
...
    accuracy                           0.99       440
   macro avg       0.99      0.99      0.99       440
weighted avg       0.99      0.99      0.99       440
```

5.4.6 XGBoost

```
XGBoost's Accuracy is:  0.9931818181818182
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       1.00      1.00      1.00        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       0.96      1.00      0.98        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       1.00      0.93      0.96        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       0.96      1.00      0.98        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
   mothbeans       1.00      0.95      0.97        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      1.00      1.00        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       0.94      1.00      0.97        16
...
    accuracy                           0.99       440
   macro avg       0.99      0.99      0.99       440
weighted avg       0.99      0.99      0.99       440
```

5.4.7 Accuracy Comparison



6 Performance Test

```
data = np.array([[104,18, 30, 23.603016, 60.3, 6.7, 140.91]])
prediction = RF.predict(data)
print(prediction)
```

```
['coffee']
```

```
data = np.array([[83, 45, 60, 28, 70.3, 7.0, 150.9]])
prediction = RF.predict(data)
print(prediction)
```

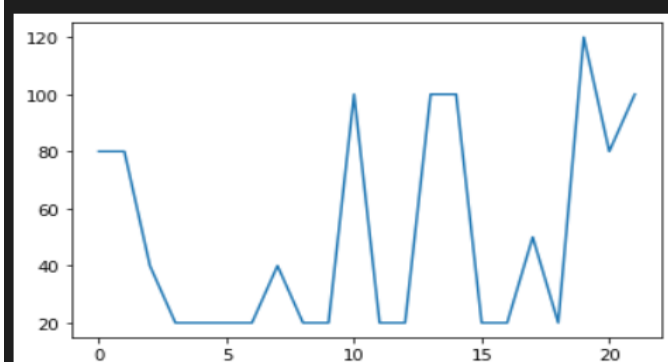
```
['jute']
```

6.1 Test Plan/ Test Cases

6.2 Test Procedure

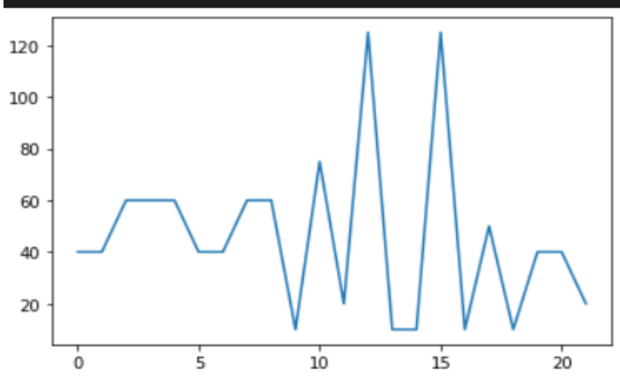
```
plt.plot(merge_fert["N"])
```

```
[<matplotlib.lines.Line2D at 0x294c8c8bfa0>]
```



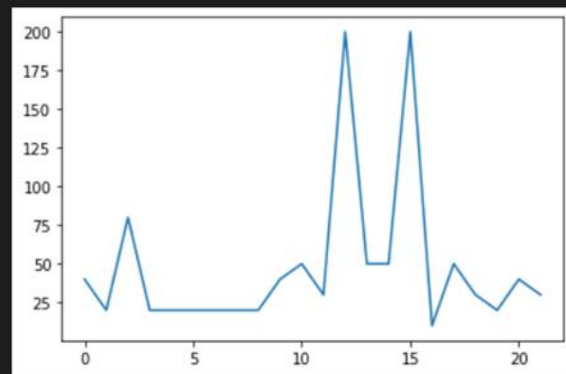
```python
plt.plot(merge_fert["P"])
```

```
[<matplotlib.lines.Line2D at 0x294cad60550>]
```




```python
plt.plot(merge_fert["K"])
```

```
[<matplotlib.lines.Line2D at 0x294cadb7b50>]
```




```python
sns.heatmap(merge_fert.corr(),annot=True)
```

```
<AxesSubplot:>
```

6.3 Performance Outcome

```
# Checking if everything went fine
df = pd.read_csv('../Data-processed/crop_recommendation.csv')
```

```
df.head()
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

```
df.shape
```

```
(2200, 8)
```

7 My learnings

- Have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
- Have an understanding of the strengths and weaknesses of many popular machine learning approaches.
- Appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
- Be able to design and implement various machine learning algorithms in a range of real-world applications.
- Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies
- Ability to understand and apply scaling up machine learning techniques and associated computing techniques and technologies.

8 Future work scope

- User Personalization: Implement personalized user experiences through AI-driven recommendations and customization options.
- Security and Compliance: Prioritize security and compliance features that were deferred due to time constraints. Stay updated with the latest security threats and industry regulations to ensure the product remains robust and compliant.
- Advanced Features: Explore the development of advanced features or functionalities that were initially out of scope but could enhance the product's value. This might include integrating AI or machine learning for predictive analytics, natural language processing for better user interactions, or augmented reality features.
- Enhanced User Experience: Invest in user experience research and design to improve the overall usability and user satisfaction of the product. This could involve conducting usability testing, implementing user feedback, and fine-tuning the user interface.
- Scalability: If the project was limited in terms of scalability, plan for future enhancements that allow the product to handle larger user bases or more extensive data volumes. This might involve redesigning the architecture, adopting cloud-based solutions, or optimizing performance.
- Integration with Third-party Services: Explore partnerships and integrations with other complementary services or platforms that could enhance the product's functionality and value.
- Data Analytics and Reporting: Develop comprehensive data analytics and reporting capabilities to provide users with valuable insights and help them make data-driven decisions.
- Community Building: Establish a community around the product to foster user engagement, gather feedback, and build a user-driven development roadmap.
- Documentation and Training: Invest in thorough documentation and training materials to make it easier for users to understand and utilize the product effectively.
- Performance Optimization: Continuously work on optimizing the product's performance to ensure it runs smoothly even as it scales.