**Team: Devashish Mahajan**

Team Members:

1) Devashish Mahajan

2) Rohit Chaudhari

**Steps:**

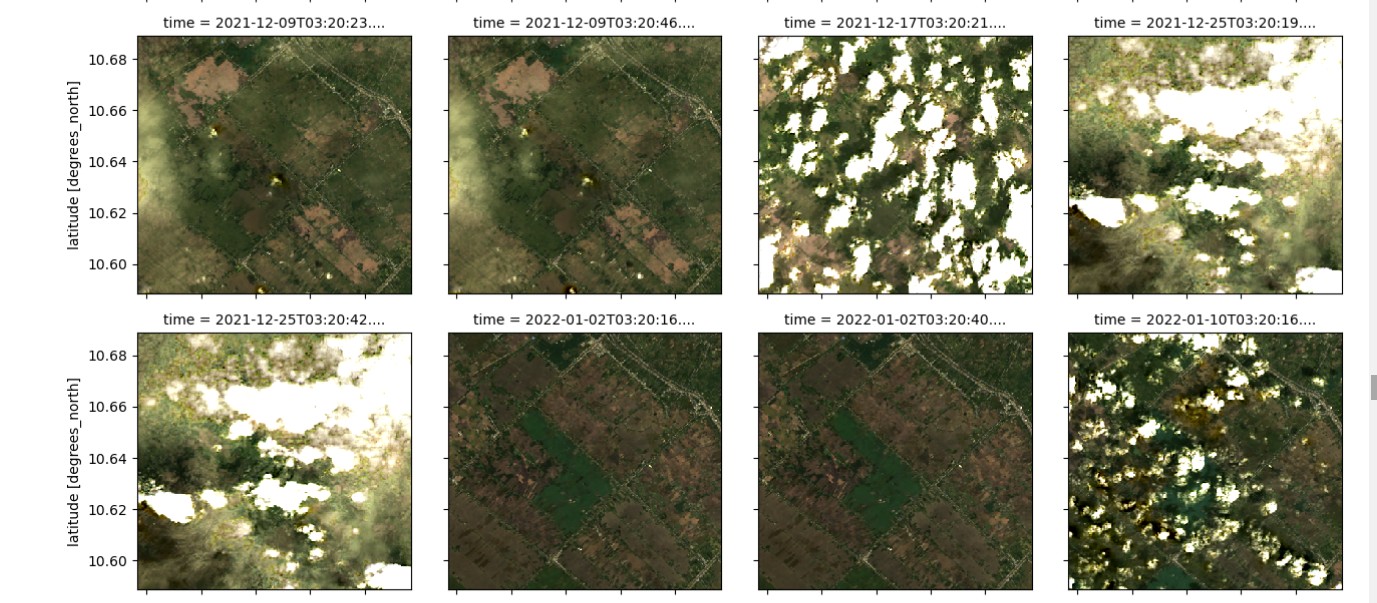
1. Data gathering
2. Data visualization (pictures and graphs)
3. Data gathering on proper hyper parameters (box size, time window)
4. Feature extraction and feature engineering
5. Model hyperparameters tunning
6. submission
7. **Data gathering:**

+ Crop\_Location\_Data\_20221201.csv --- Document provided for training data

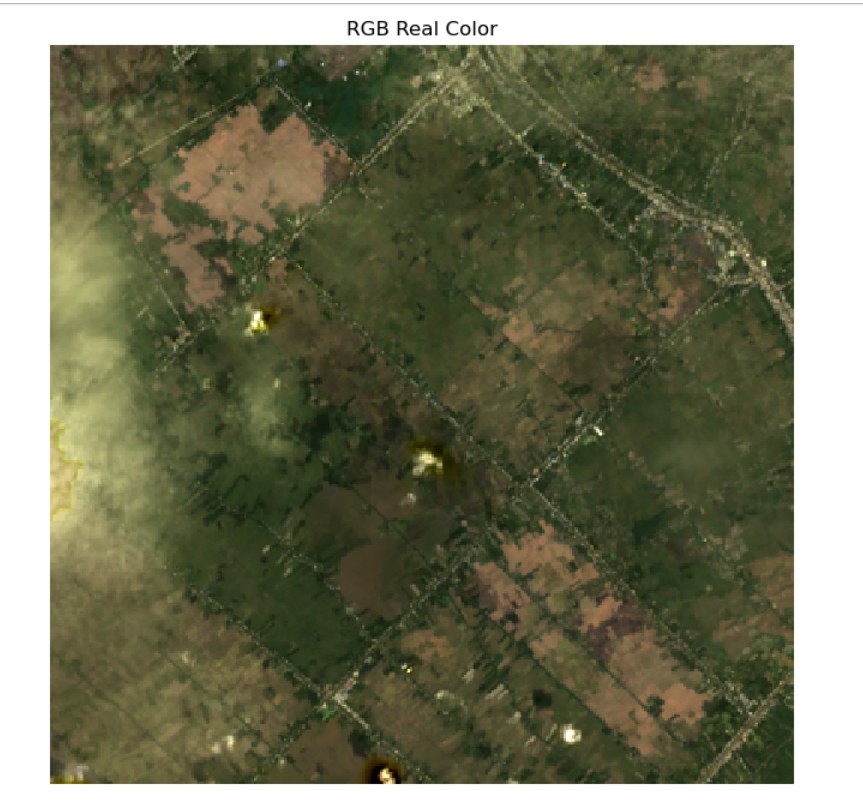
1. **Data visualization (pictures and graphs):**

+ Hyperparameter tuning for Data Gathering From Landsat.ipynb

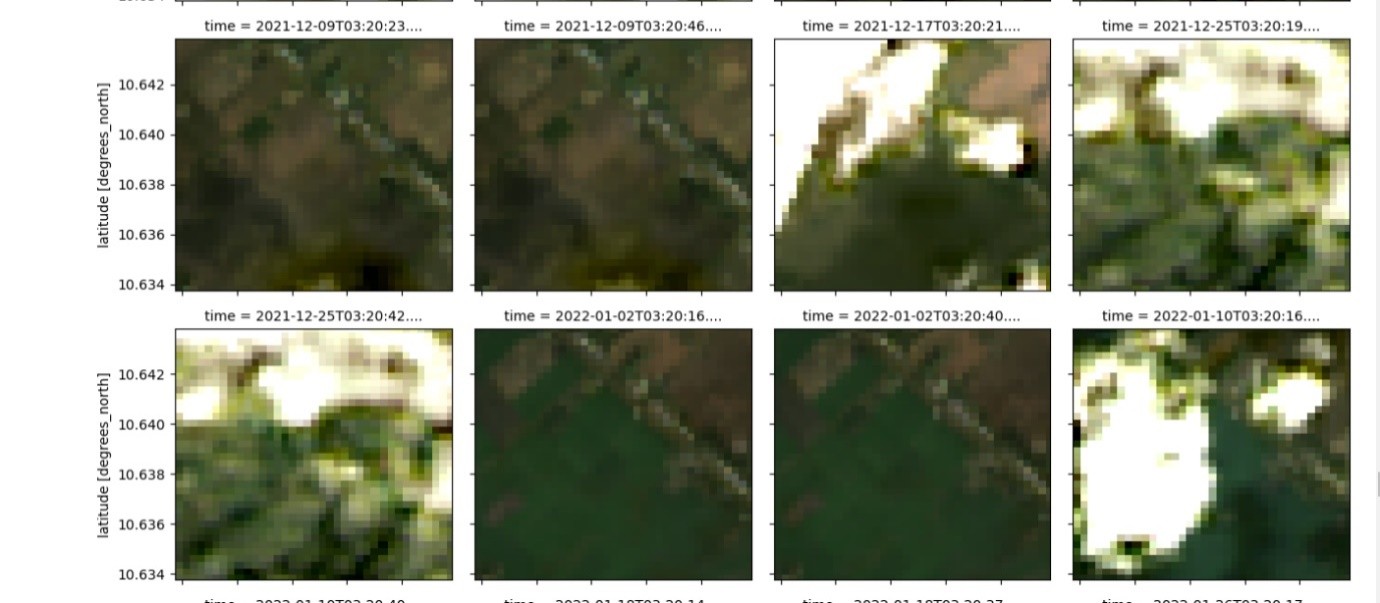
* Tunning hyperparameter "box\_size\_deg" to enclose only required area around the logitude and latitude location. (initial=0.10, final=0.010)



**Figure 1.1 Satellite images for box\_size\_deg = 0.1**



**Figure 1.2 Satellite images for box\_size\_deg = 0.1 and area = 123.21 sq. km**.

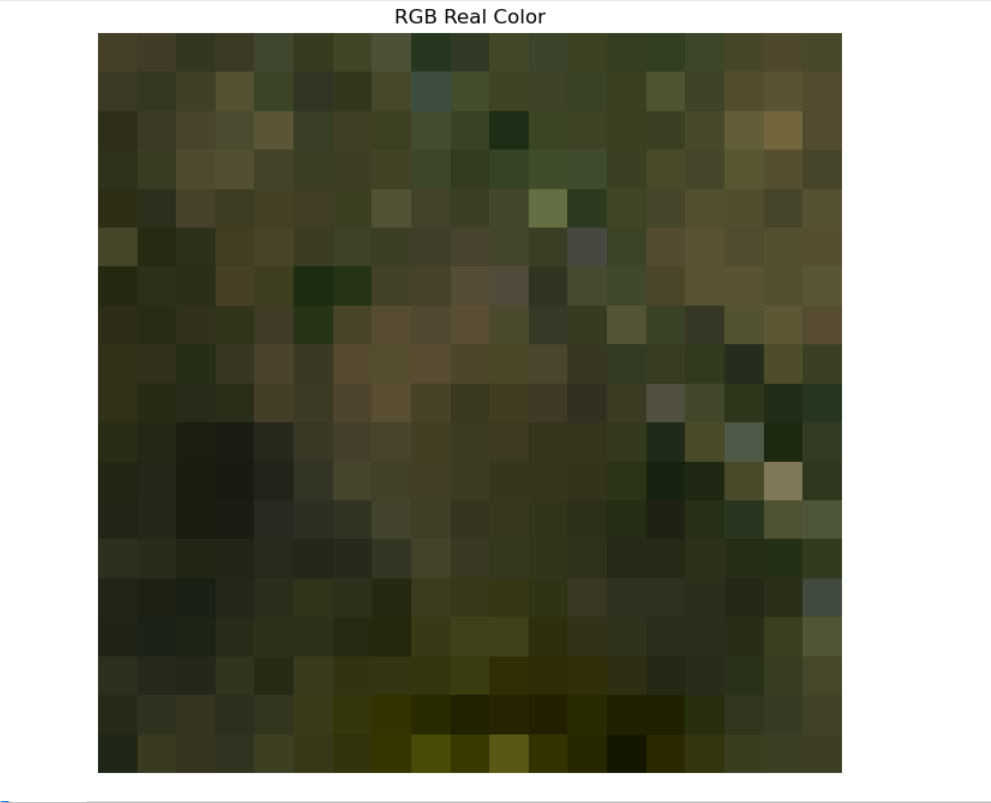


**Figure 1.3 Satellite images for box\_size\_deg = 0.01**

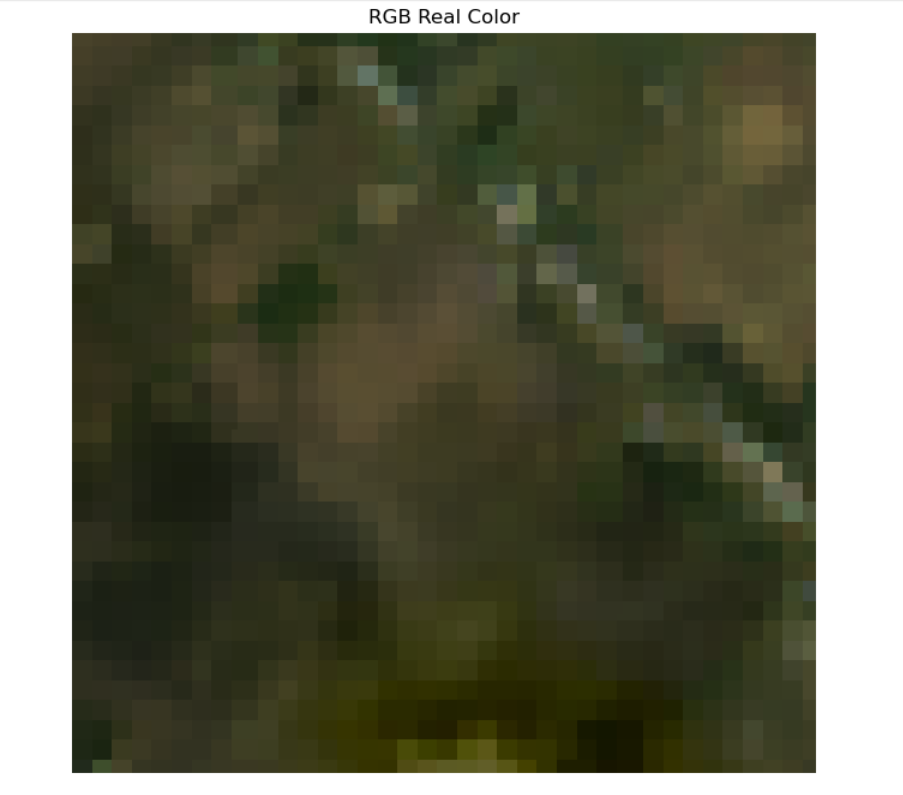
* Tunning hyperparameter "time\_window" to cover the crop cycles.

(final ="2021-11-01/2022-05-30")

* Tunning hyperparameter "resolution" to get good quality image resolution. (initial=60, final=10)

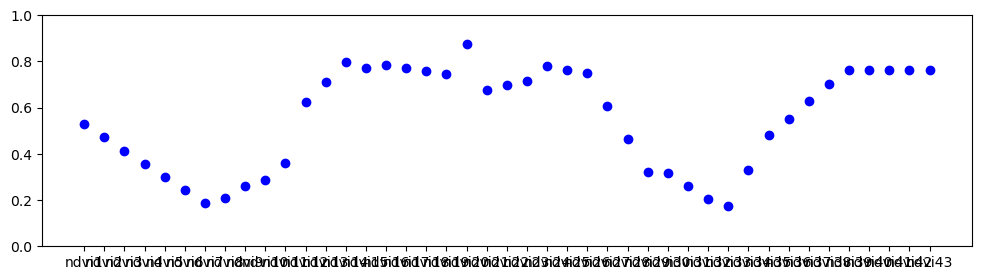


**Figure 1.4 Satellite images for resolution = 60**

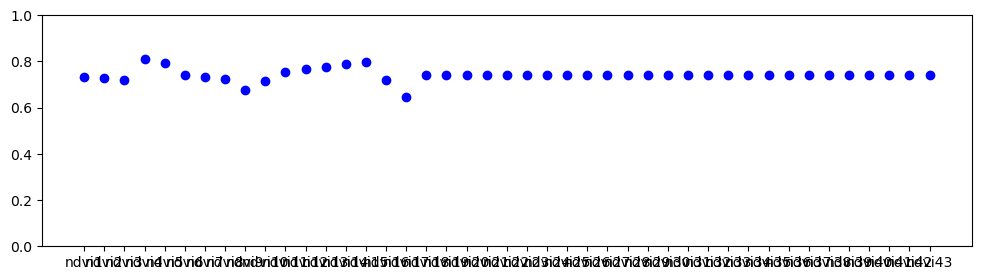


**Figure 1.5 Satellite images for resolution = 10**

* Study the graph patterns generated from NDVI for Rice.

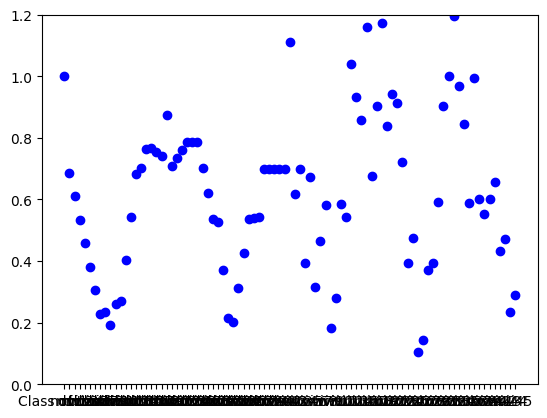


**Figure 1.6 NDVI Scatter Plot for Rice**



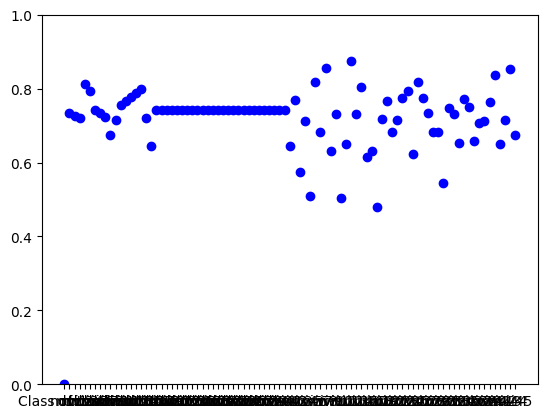
**Figure 1.7 NDVI Scatter Plot for Not Rice**

* Presence of rice shows hills and valleys in the scatter plot.



**Figure 1.8 NDVI combined RVI Scatter Plot for Rice**

* Non rice region (river, road, Jungle) shows smooth and flattened NDVI scatter plot.



**Figure 1.9 NDVI combined RVI Scatter Plot for Not Rice**

1. **Data gathering on proper hyper parameters (box size, time window):**

**Jupyter Notebook**: RVI Data Gathering From Sentinel 1.ipynb

RVI parameters:

* satellite= "sentinel-1-rtc"
* box\_size\_deg= 0.0004
* time\_window= "2021-12-01/2022-08-30"
* resolution= 10

**Jupyter Notebook**: NDVI Data Gathering From Landsat.ipynb

NDVI parameters:

* satellite= "landsat-8", "landsat-9"
* box\_size\_deg=0.010
* time\_window="2021-11-01/2022-05-30"
* resolution=10

1. **Feature extraction and feature engineering:**

+ Imputing missing values:

>>>vh\_vv\_data.interpolate(method='linear',axis=1, inplace=True)

>>>vh\_vv\_data.interpolate(method='linear',limit\_direction='backward',axis=1, inplace=True)

1. **Model Training And Features Selection:**

**Jupyter Notebook**: Model Training And Features Selection-Bagging With Voting.ipynb

## Notebook contains following sections:

1. Loading Train Data And Test Data
2. Function to extract features from data
3. Check for correlation using Heatmap
4. Check Distribution of Train Data and Test Data using Scatterplot
5. Check Distribution of Train Data and Test Data using Distplot
6. 2 sample Kolmogorov–Smirnov test
7. Results of 2 sample Kolmogorov–Smirnov test
8. Check for Variance Inflation Factor (VIF)
9. Check the distribution of train data and test data for all exctracted features
10. Model
11. Grid Search for Hyperparameter Tuning
12. Grid Search Results
13. Final model after hyperparameter tuning
14. Save Model
15. Make Predictions on Test Data and Save Prediction file
16. **submission:**

+ submission\_clf\_final\_20230330.csv