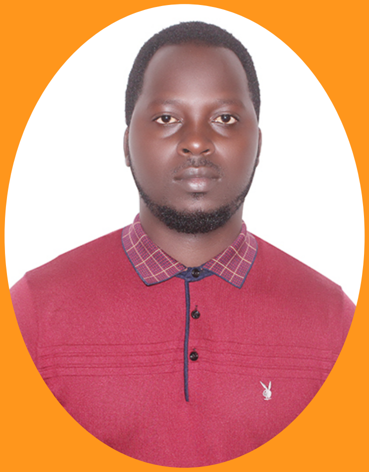
Harnessing Machine Learning for Crop Yield Prediction in Rwanda.

Unleashing the Power of Landsat 8 Satelite Imagery and Seasonal Agriculture Survey.

11 June, 2023

UNIVERSITY OF RWANDA

About Facilitator



Murera Gisa

Senior Data Scientist

National Bank of Rwanda (BNR), AIRA Analytics(Uganda)
Home-based Consultancy

@myblog @gisa_murera @mgisa 0788266517 elgisamur@gmail.com

About the Workshop:

In this workshop we hope to demonstrate the application of ML in agriculture sector. We want to upskill all participants in order for them to understand the combination of different data sources for optimal ML prediction.

Remote Sensing Section: Show the participants how to access the satellite imagery through the GEE and process workflow.

Machine Learning: Explain them the relevant theorem in ML (Non Free Lunch Theorem, NFLT)and apply it on the treated data set for crop yield prediction.

Goal:

- Upskill all participants to understand GEE code Editor and Climate Data Tool embedded in R used in the remote sensed climate data pipeline

> The training only aims to serve as a foundation for participants' ML career journey.

- Familiarisation with Supervised Machine Learning pipeline

> Upskill all participants to understand the protocol for Supervised Machine Learning Techniques.

Objectives:

Explore the different data sources relevant to the ML project.

Validate the data collected to achieve the accurate ML results. Develop a robust machine learning model.

Assess the predictive accuracy of the model.

Key outcomes:

At the end of this workshop, participants will be able to:

> Have a basic understanding of the GEE Code Editor and remote sensed data processing pipeline.
>
> Understand the different way to collect the weather and climate data using the magic potential of R.
>
> Be able to do a basic exploratory analysis on GEE
>
> unleash the power of R software to robust train ML Models.
>
> Understand NFLT and Assess the performance of models
>
> Take predictive analysis to the next level through automation and presentation in R {modeldown} library.

Asking assistance

> PLZ!!! Ask questions, we've been down this road before!
>
> Please feel free to stop me and ask a question.
>
> If you feel more comfortable asking questions in writing feel free to email them to `elgisamur@gmail.com`
>
> Help each other out! Some might be further along their data journeys than others.

Summary:

Have a basic understanding of the Climate Science, GEE and R code used in a data pipeline.

Understand the flow of predictive analysis pipeline.

Being able to do a basic data exploration and traitment for predictive analysis.

Visualize ML results using {ggplot2}.

Create interactive visual dashboard for ML results.

Table of content

- Hour1: Setting of working environment and Introduction to the project background
- Hour2: Interacting with GEE and CDT for Data collection, Visualization and Descriptive analysis
- Hour3: Introduction to NFLT and Predictive Analysis pipeline

Introduction and Project Background

- As the world's population continues to grow at an unprecedented rate, as climate keep changing unexpectedly, ensuring global food security becomes an increasingly pressing challenge.
- The agriculture, being the backbone of our sustenance, and contribute to the global economy movement.
- In Rwanda, agriculture plays a vital role in the country's economy, contributing to nearly a third of its GDP.
- This requires innovative technological solutions to maximize productivity and optimize resource utilization.
- This is where Artificial Intelligence and Machine Learning steps in.

Problem Statement

- Despite the importance of crop yield prediction, achieving optimal and trustworthy accuracy remains a challenge due to the complex interactions between weather, soil characteristics, climate factors, crop health, and other agricultural practices.
- The use of traditional and classical methods of crop yield estimation failed (slowness, poor parameter tuning, poor computation, and resource-intensive) to analyze large crop data set from different sources including images, hence poor estimation of crop yield and often requires experimental judgments.

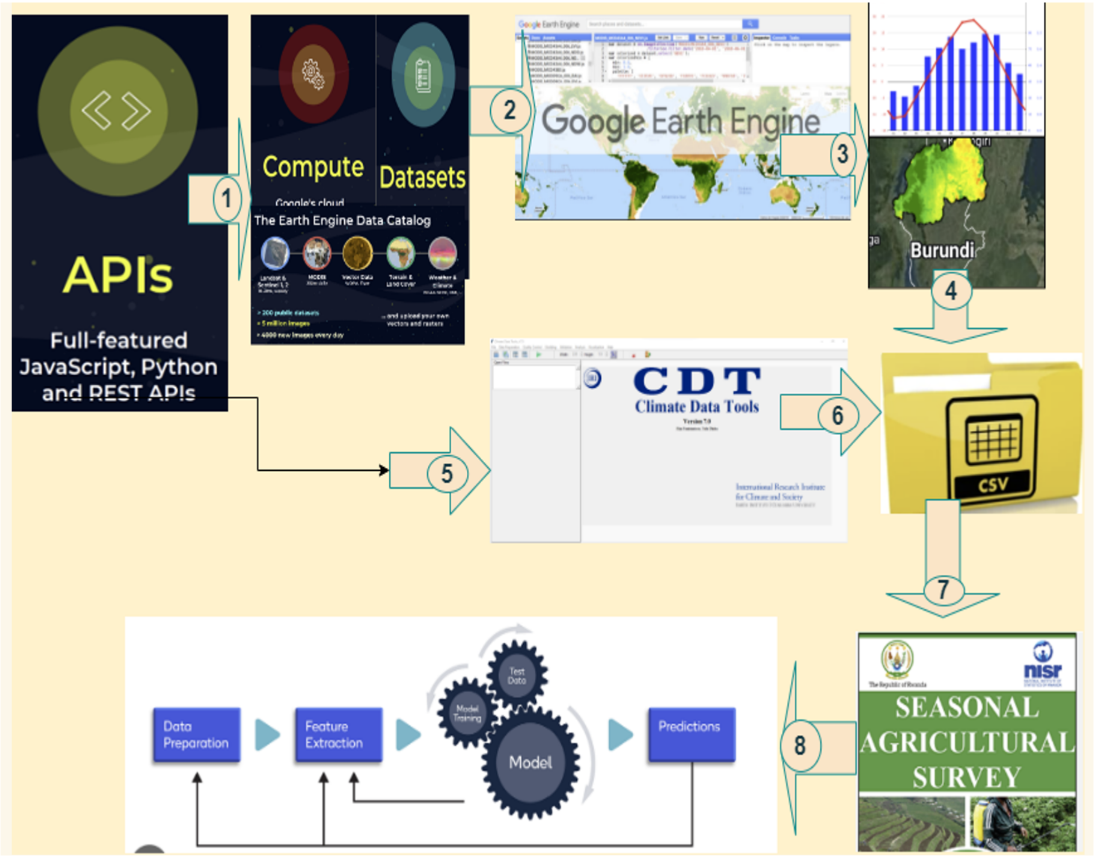
- However, recent advances in the technological design of satellites and drones with high-resolution lenses and multiple bands combined with Artificial Intelligence(computer vision and deep Learning) could offer a promising opportunity to highly improve crop yield prediction for evidence-based decision-making.
- Supervised Machine Learning algorithms incorporate various data sourced from NASA, UN-SPIDER Data Portal, TAMSAT, JRA-55, and NISR to accurately predict crop yields in Rwanda and provide farmers and policymakers with critical information to make informed decisions about their agricultural practices and inputs.

Relevancy of the project and its Scope

- Crop yield is a crucial metric in agriculture, as it determines the productivity and profitability of a farm.
- It is defined as the amount of crops produced per unit area of land
- It is influenced by a range of factors including soil fertility, weather conditions, the type of crop grown, pest and disease control, and agriculture practices and management.
- The accurate prediction of crop yield is critical for farmers, stakeholders, and policymakers to make informed decisions about agricultural production, food security, and food price in Rwanda.
- The project results will also be used to optimize crop management practices and ultimately improve crop yield in the future, hence food security happens.

Methodological Workflow

Method Flow


APIs
Full-featured JavaScript, Python and REST APIs
Compute
Datasets
The Earth Engine Data Catalog
Google Earth Engine
Burundi
CDT
Climate Data Tools
CSV
SEASONAL AGRICULTURAL SURVEY
nisr
Data Preparation
Feature Extraction
Model Training
Test Data
Model
Predictions
1
2
3
4
5
6
7
8

Steps

- Steps 1&5: APIs creation to access the NASA,COPERNICUS, TAMSAT, and JRA-55 geo-spatial data repositories respectively.
- Steps 2&3: Accessing GEE super computational facility to explore and visualize the imagery data collected by LandSat (30X30 m resolution and once every 2 weeks)
- Steps 4&6: Uploading Rwanda border shapefile indicating the mapped station to the visual presentation of data and numeric downloading for predictive analysis.
- Step 7: Merging the pre-processed and cleaned satellite data (weather and soil) with crop information for supervised predictive analysis.
- Step 8: Predictive Regression analysis

Data Sources

We limited ourselves to the LANDSAT-8 Imagery Data stored from different data repositories

- Through GEE Code Editor we access remotely Sensed data from:

```
- NASA
- Copernicus and Landsat Program
- UN-SPIDER Data Portal
```

- TAMSAT: Tropical Applications of Meteorology using Satellite data and ground-based observations (Blended data).
- JRA-55/JMA: Japanese 55-year Reanalysis of Meteorological data.
- NISR (SAS2018-SAS 2022): National Institute for Statistics of Rwanda

Why LandSat-8 ?

- Spatial resolution: 30X30m
- Temporal resolution: 16-day repeat cycle (overpassed every 16 days and it crosses the Equator every 8 days).
- Powerful satellite for regularly monitoring global change.
- Latest launched in 2013 and its data is public
- Spatial Orbit: Stratosphere: 705km from the ground

Data Description

year	seasons	districts	crop_type	production(T)	arable_size(Ha)	air_temp(oC)	atm_pressure(Pa)	evapo_trans(Kg/sq.m)
2018	Season A	Bugesera	banana	31707.35	4741.34	22.76	782.25	0.06
2018	Season A	Bugesera	banana	5171.67	1685.41	22.76	782.25	0.06
2018	Season A	Bugesera	banana	33314.41	6560.46	22.76	782.25	0.06
2019	Season A	Bugesera	banana	71692	12881	21.86	784.17	0.12
2020	Season A	Bugesera	banana	21892.31	4455.2	13.01	776.83	0.13

humidity(g/m3)	l_stemp(oC)	ndvi	rainfall (mm)	soil_moisture(Kg/m3)	solar_rad(Kw/m2)	wind_force(Pa/m2)
57.01	24.37	0.16	86.28	0.31	339.72	-0.86
57.01	24.37	0.16	86.28	0.31	339.72	-0.86
57.01	24.37	0.16	86.28	0.31	339.72	-0.86
65.41	23.22	0.14	118.8	0.49	340.76	-0.61
61.13	23.8	0.12	88.3	0.4	340.07	-0.66

Practical Session

Requirements:

> Stable internet to access the online data repositories and easy packages installation
>
> Installed R and Rstudio IDE

Booooooo!!!! Let's code together to apply ML technique to inform policy makers for sustaining food stability in Rwanda.

THANKS