Earth Observations and ML for estimating agricultural activity to Support the Planning and Execution of Agricultural Census Events.

Laboratory of Data Science

General Directorate of Integration, Analysis, and Research

In collaboration with:

- **Regional Directorate of the North-Central Region (DRCN)**
- **General Directorate of Economic Statistics (DGEE)**

November 2024

Introduction

Introduction

- INEGI, as part of its duties, is responsible for generating the statistical and geographical information that the country demands to support decision-making and the definition of public policy.

- Remote sensing is an important tool in the study of natural resources and the environment, where it's possible applications are many, with one of the most important being agricultural monitoring.

- The application of new technological trends presents an excellent scenario to support the generation of information, leading to the integration of innovations for process improvement.

Introduction

- With the experience gained from events such as the 2022 Agricultural Census, several techniques were applied:
  - The identification of cultivated and uncultivated agricultural areas.
  - The determination of irrigated agricultural zones.
  - Understanding the conditions of agricultural areas at different stages.
  - The identification of crops.
- It enabled the identification of areas of opportunity to support activities focused on the better use of these technologies for the generation of information for the sector.

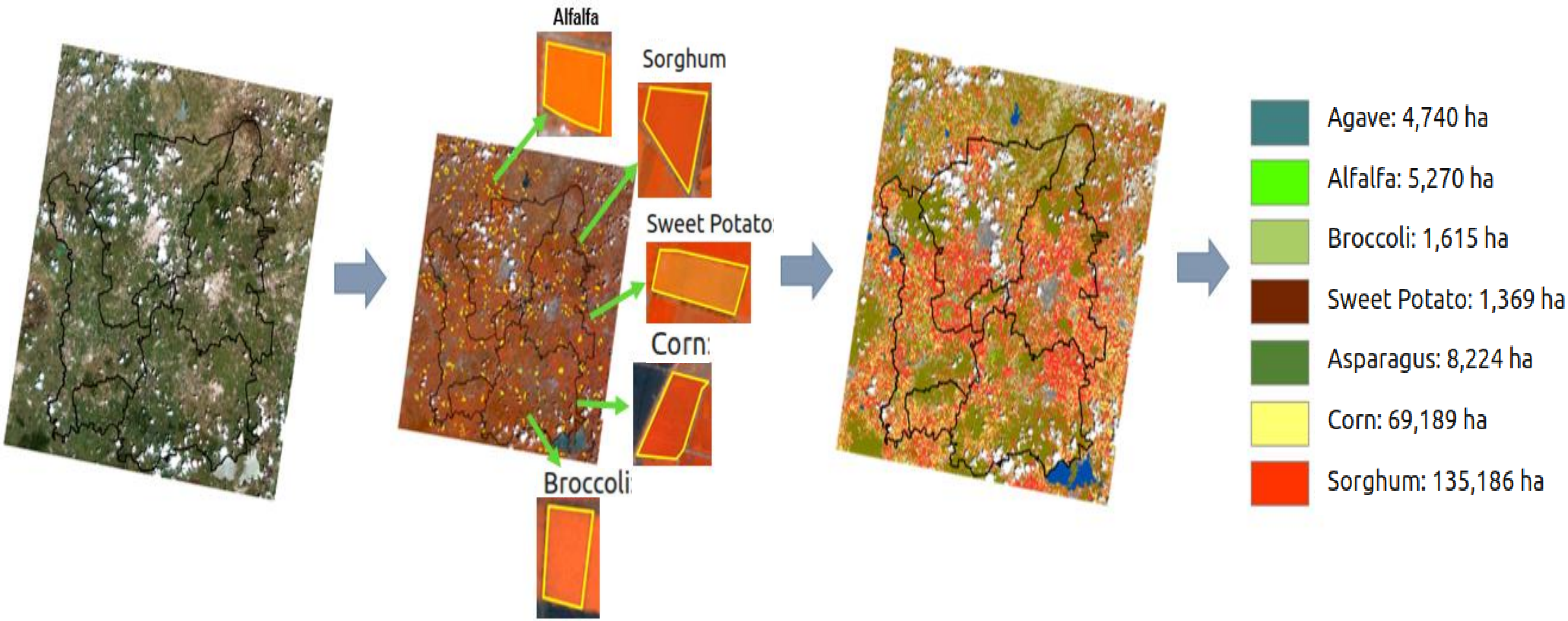
Objectives

- Define a methodological procedure for the analysis of agricultural activity, through cloud processing of satellite images.
- Detect agricultural activity through multitemporal analysis of vegetation indices from satellite images.
- Identify areas with irrigation modality, through multitemporal analysis of vegetation indices from satellite images.
- Identify crops of interest and estimate areas for validation or comparison with historical data.

Prior Experiences in Crop Identification Using Satellite Images (DGEE)

INEGI

Methodology



It has been applied to different crops and regions of the country

Stage 1
Experiences 2013-2016

Purchased staellite images (spatial resolution 6.5 meters)

Software **comercial**

Field samples

- Obtained through field visits to selected areas

Achievements

- Good definition and positive results for some crops
- High accuracy and sufficiency in the field samples

Stage 2
Experiences 2018-2020

Free satellite images (spatial resolution 10 meters)

Open-source software

Field samples

- Obtained through **administrative agricultural records** (validated in the office).

Achievements

- Free software and images
- Although limited, administrative records were obtained for free Good definition and positive results for some crops

RapidEye idrisi

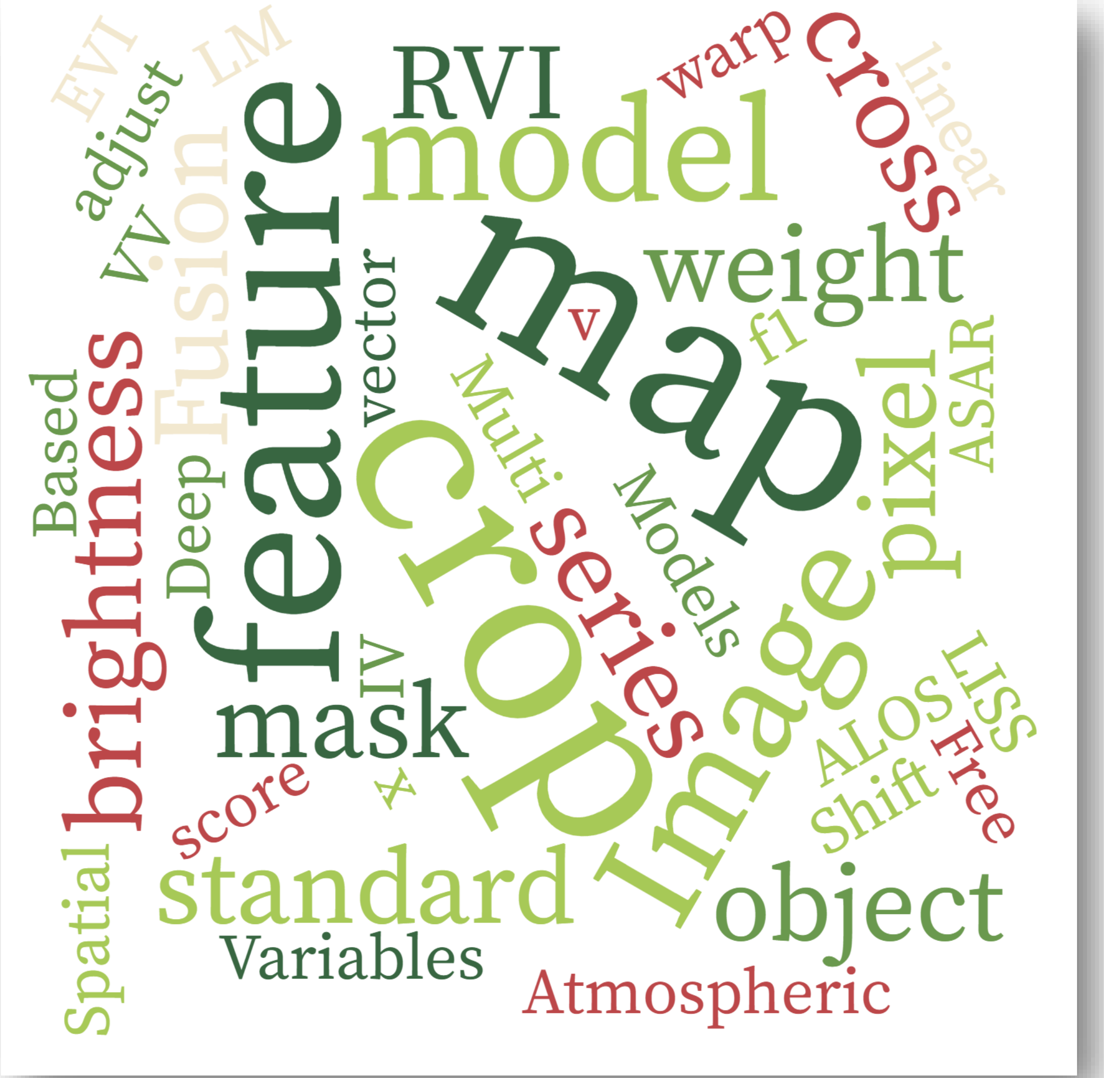
sentinel-2 QGIS Orfeo ToolBox

What We Have Not Used

- Time series.
- Radar images such as Sentinel-1.
- Segmentation.
- Additional layers such as Digital Elevation Models, climate, spectral indices, among others.
- Geomedian (representation of a set of cloud-free images).

Research Project Progress – 2024

Literature Review



1. **Defourny, P., et al.** "Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world." *Remote Sensing of Environment* 221 (2019): 551-568.

2. **Belgiu, M., & Csillik, O.** "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis." *Remote Sensing of Environment* 204 (2018): 509-523.

3. **Estes, L.D., et al.** "High resolution, annual maps of the characteristics of smallholder-dominated croplands at national scales." *EarthArXiv* (2021).

4. **Valero, S., et al.** "Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions." *Remote Sensing* 8.1 (2016): 55.

5. **Cheng, G., et al.** "Crop type classification with combined spectral, texture, and radar features of time-series Sentinel-1 and Sentinel-2 data." *International Journal of Remote Sensing* 44.4 (2023): 1215-1237.

6. **Xiong, J., et al.** "Automated cropland mapping of continental Africa using Google Earth Engine cloud computing." *ISPRS Journal of Photogrammetry and Remote Sensing* 126 (2017): 225-244.

7. **Vorobiova, N.S.** "Crops identification by using satellite images and algorithm for calculating estimates." *CEUR Workshop Proceedings* 1638 (2016).

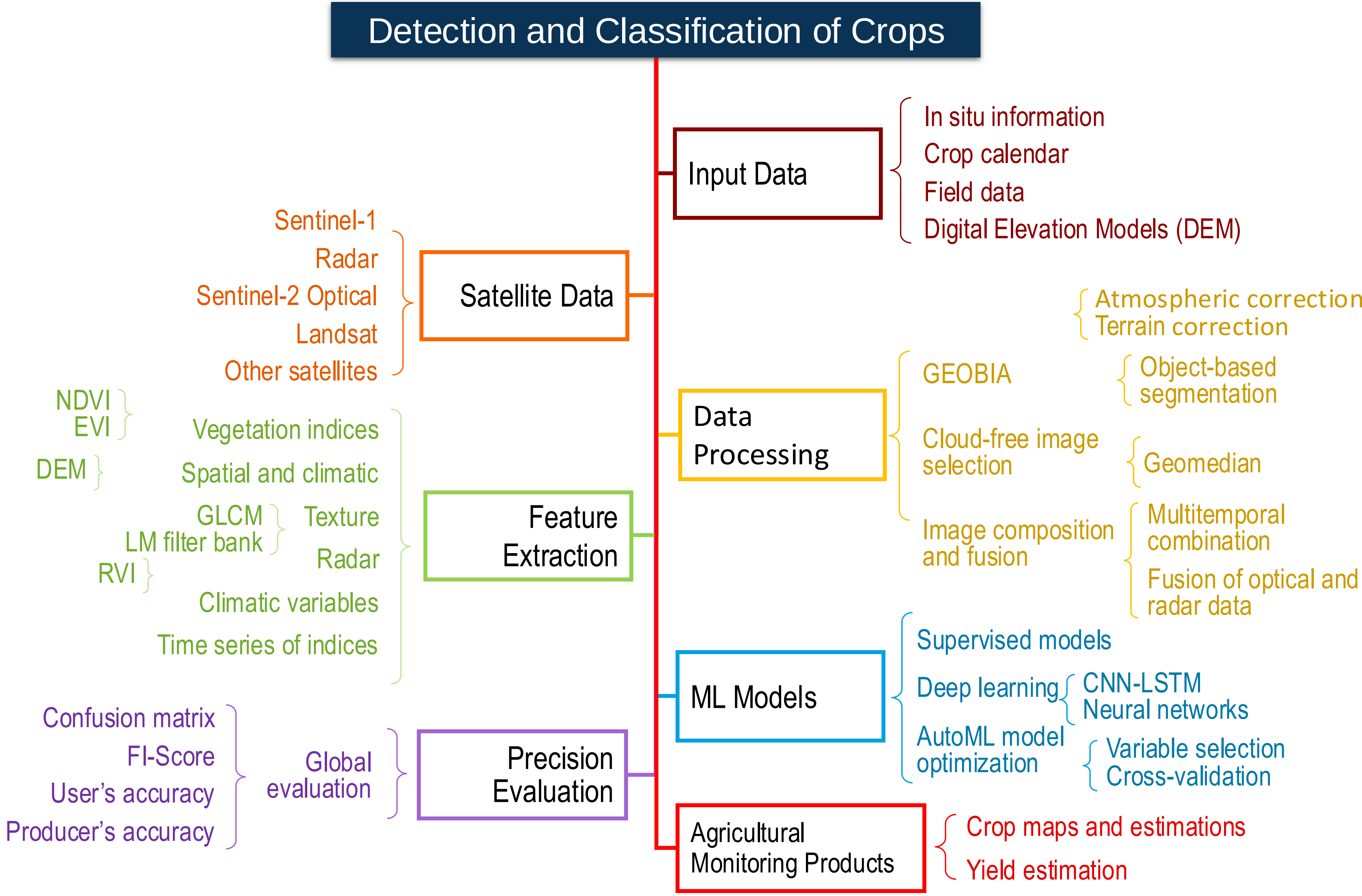
8. **Blickensdörfer, L., et al.** "Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2, and Landsat 8 data for Germany." *Remote Sensing of Environment* 269 (2022): 112831.

9. **Bahrami, H., et al.** "A Meta-Analysis of Remote Sensing Technologies and Methodologies for Crop Characterization." *Remote Sensing* 14.22 (2022): 5633.

10. **Alami Machichi, M., et al.** "Crop mapping using supervised machine learning and deep learning: A systematic literature review." *International Journal of Remote Sensing* 44.8 (2023): 2717-2753.

11. **Snevajs, H., et al.** "Crop Detection Using Time Series of Sentinel-2 and Sentinel-1 and Existing Land Parcel Information Systems." *Remote Sensing* 14 (2022): 1095.

Thematic delimitation for crop detection and classifications



Test Environment (Sandbox)

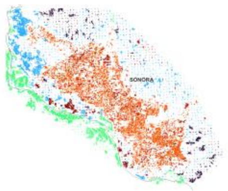
- Cluster with 160 cores, 1.5 TB RAM, 16 TB storage
- Preproduction and training environment
- JupyterHub, Docker Swarm
- Python on Ubuntu
- Grid Storage RAID (Data Lake)
- 45 TB NAS
- On-premise infrastructure

INEGI

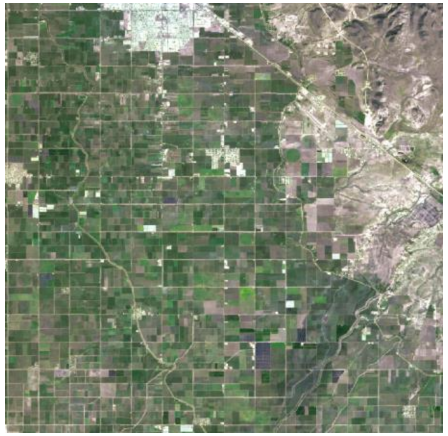
Test Environment (Sandbox)

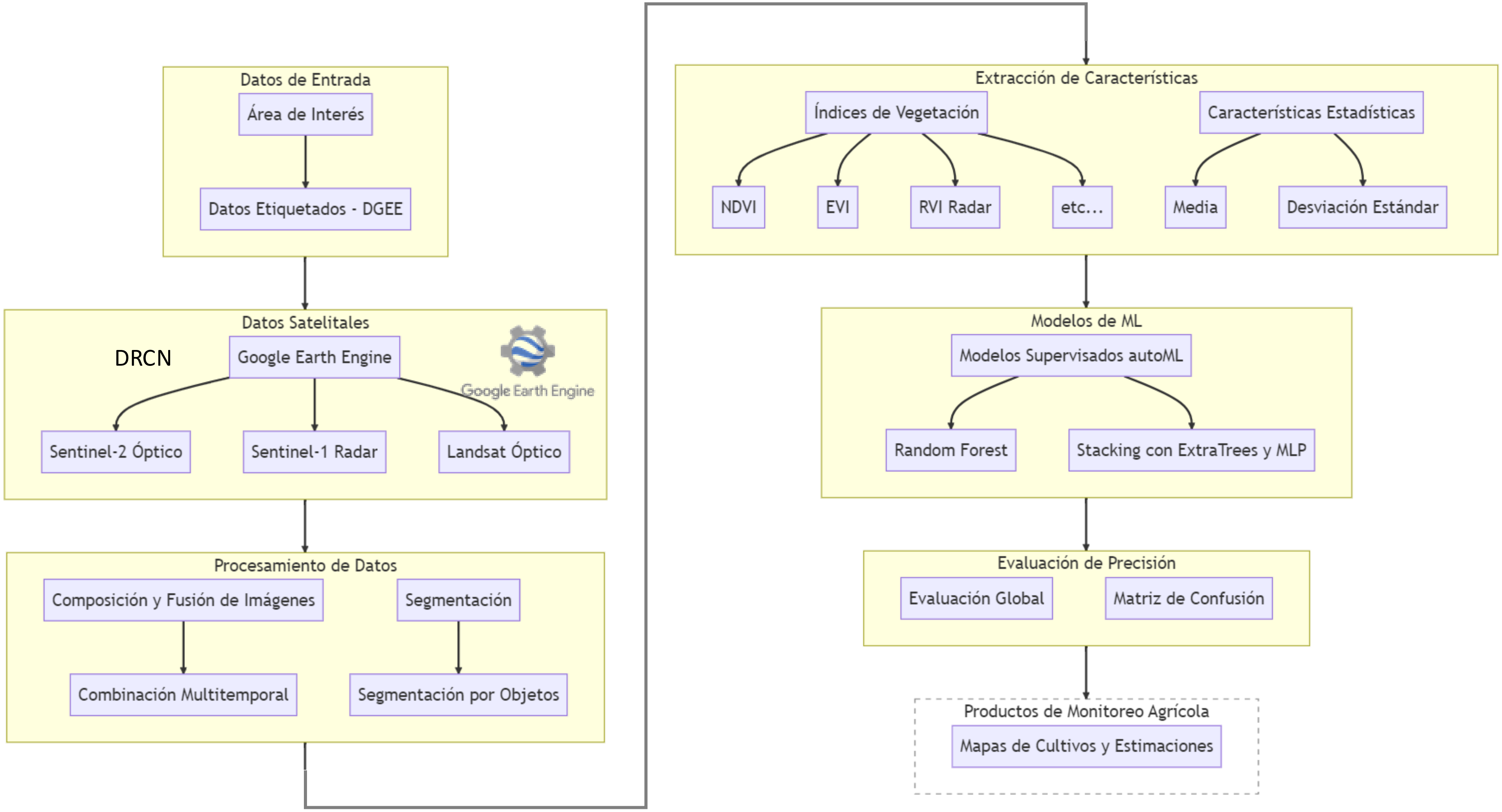
INEGI

Crop Classification in Python



Sonora: Valle del Yaqui (DGEE)
Temporada: Otoño – Invierno
2018-2019





Crop Classification in Python (Sandbox)

Tutorial_Satellite.ipynb

Code

No Kernel

TIFF comprimido sin pérdida. Esta función es particularmente útil cuando se trabaja con grandes áreas de estudio que han sido divididas en partes más pequeñas para su procesamiento. La función `create_unified_tif` automatiza el proceso de unificación, asegurando que el resultado final sea un archivo TIFF consolidado, listo para su análisis.

[15]:

```
def create_unified_tif(input_pattern, output_tif, compression='LZW'):
    """
    Crea un archivo TIFF unificado a partir de fragmentos utilizando `gdal_merge.py`.

    :param input_pattern: Ruta y patrón de archivos TIFF fragmentados (ej. 'shared/cultivos/GM-O-I-2017-2018/*.tif')
    :param output_tif: Nombre del archivo TIFF unificado de salida.
    :param compression: Tipo de compresión sin pérdida para el archivo TIFF (por defecto es 'LZW').
    """
    # Obtener la lista de archivos que coinciden con el patrón
    tif_files = glob.glob(input_pattern)

    if not tif_files:
        print("No se encontraron archivos TIFF que coincidan con el patrón.")
        return
```

Crop Classification in Python (Local Setup)

Crop Classification in Python: A Local Installation Tutorial with Google Earth Engine Integration.

The first step is to install the necessary environment for everything to work on our personal computer. This tutorial assumes that the user has permissions to install software on a Windows machine.

To begin, we need to install a version of Python that makes it easy to install packages. For this, we will use a tool called **Miniforge**. Miniforge is a minimal installer that sets up **conda** to use the free and open-source **conda-forge** channel by default. This simplifies the installation and management of Python and its libraries.

You can find Miniforge at the following link: [Miniforge GitHub](). When you open the

```
train_data, test_data = split_data(balanced_data)

target_class = 'land_cover'
pipeline_name = "Extra-Trees for SONORA Time Series RADAR"
fit_and_test(train_data, test_data, target_class, exported_pipeline, pipeline_name)
```

```
Inicia entrenamiento
Clasificación
El resultado de la clasificación con el pipeline: Extra-Trees para SONORA Serie de Tiempo RADAR
```

	precision	recall	f1-score	support
1	0.9202	0.9804	0.9494	153
101	0.7651	0.8759	0.8167	145
102	0.8169	0.7682	0.7918	151
103	0.7467	0.7671	0.7568	146
104	0.7143	0.5645	0.6306	62
105	0.6667	0.6809	0.6737	47
106	0.8409	0.7872	0.8132	47
107	0.5238	0.6471	0.5789	17
108	0.7000	0.4667	0.5600	15
109	0.5000	0.2857	0.3636	7
110	0.4792	0.4182	0.4466	55
accuracy			0.7716	845

Results

Classification Results

	Precision	Sensivity	F1 Score	Support
No Cultivation	0.9202	0.9804	0.9494	153
Wheat	0.7651	0.8759	0.8167	145
Corn	0.8169	0.7682	0.7918	151
Chickpea	0.7467	0.7671	0.7568	146
Safflower	0.7143	0.5645	0.6306	62
Bean	0.6667	0.6809	0.6737	47
Alfalfa	0.8409	0.7872	0.8132	47
Potato	0.5238	0.6471	0.5789	17
Walnut	0.7000	0.4667	0.5600	15
Tomato	0.5000	0.2857	0.3636	7
Others	0.4792	0.4182	0.4466	55
Accuracy			0.7716	845
Macro Average	0.6976	0.6583	0.6710	845
Weighted Average	0.7675	0.7716	0.7669	845

Next steps

Next Steps

- Strengthen our skills in Data Science.
- Continue experiments and optimize the methodology as a team.
- Expand study areas to more regions, adapting methodologies to each context.
- Broaden the analysis to more crops and seasons.

INEGI

Conclusions

Open data and cloud

Test Area Processing

Data Science Methodology

Initial Results

Questions

GRACIAS

Conociendo México

800 111 46 34
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

INEGI Informa