Hackathon
Automatic land cover mapping

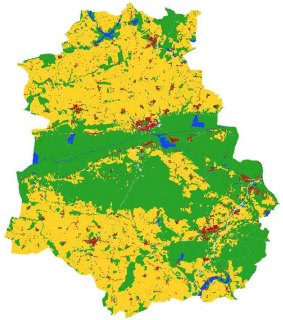
OpenGeoHub Summer School 2023

https://github.com/kadyb/OGH2023_hack
https://www.kaggle.com/competitions/ogh2023

Objective

Classification of Landsat satellite data to determine land cover classes

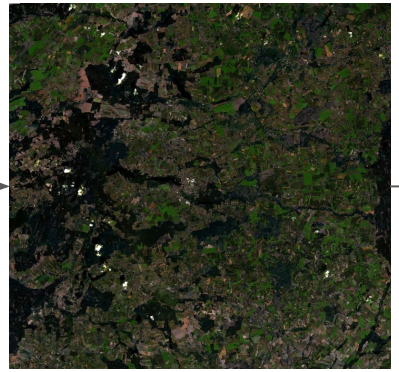
REFERENCE DATA



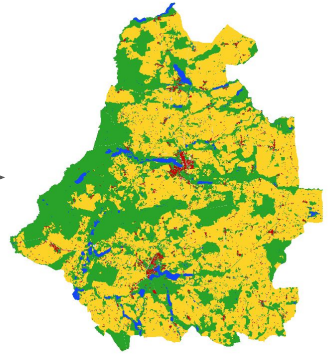
SPECTRAL BANDS

MACHINE LEARNING

NEW DATA



PREDICTION



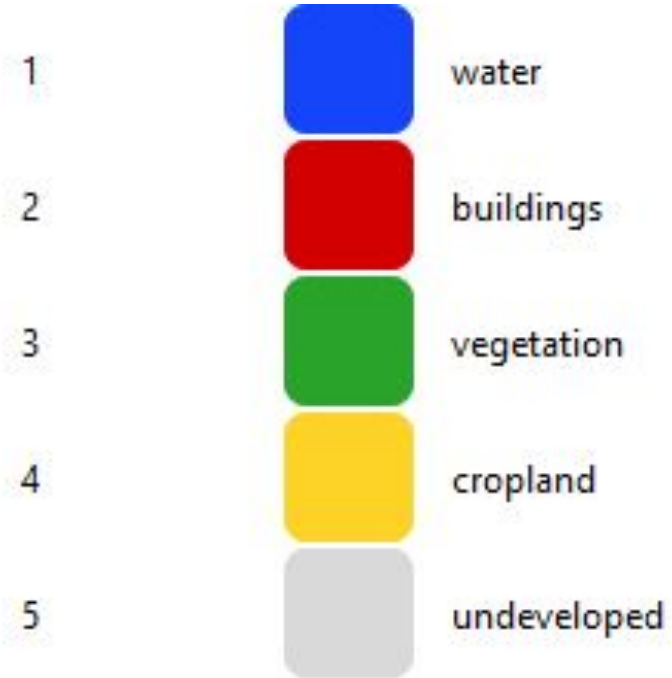
Dataset

REFERENCE DATASET:

- 5 land cover classes
- training and validation sets

EXPLANATORY VARIABLES:

- 7 spectral bands (30 m)
- panchromatic band (15 m)



Dataset is available on Kaggle.
The description of spectral bands can be found here.

Rules

1. You can submit the results individually or in pairs
2. You must be registered on Kaggle
3. Finally, you must send the results as a .csv file and a reproducible notebook
4. If you agree, we can make your notebook public in the hackathon repository
5. Maximum of three submissions per day are allowed
6. Three notebooks with the highest validation metric are subject to final evaluation. The best work is selected by the jury

Tutorial and submission

Step-by-step tutorial in R:

https://kadyb.github.io/OGH2023_hack/Submission.html

Submit the results on Kaggle:

https://www.kaggle.com/competitions/ogh2023/submissions

DEADLINE:
FRIDAY (1ST SEPTEMBER) 13:00

How to improve the result?

1. Model hyperparameter tuning
2. Ensembling different models
3. Dataset balancing (you can also combine training and validation datasets)
4. Feature engineering including:
    - Clustering
    - Tasseled cap transformation
    - Spectral indices
    - Using textural features (e.g. SAGA GIS)
    - Reducing spatial resolution
5. Additional features:
    - Radar data from Sentinel 1
    - Normalized digital surface model (nDSM)
    - Panchromatic band (15 m)
    - Satellite scene from another date(s)
6. Post-processing:
    - Modal filter
    - Sieve filter

Notes

1. Set the randomness seed in the script
2. Make sure your model returns all land cover categories
3. Make sure you use the same coordinate reference systems:
    - Polish National Geodetic Coordinate System 1992 (EPSG:2180)
    - Universal Transverse Mercator Coordinate System Zone 34N (EPSG:32634)
4. State of the art classification algorithms (generally pixel-based):
    - random forest
    - gradient boosting
    - support vector machine
5. Do not waste time on tuning the model that improves the score by 0.00001; focus more on feature engineering
6. If you have any questions (or problem) about the hackathon, feel free to ask on the hackathon channel on Mattermost

Good luck!