|  |  |
| --- | --- |
| **MODULE 6: PREDICTIVE MODELLING USING EO DATA** | |
| **OBJECTIVES** | * Understand common use cases (applications) for predictive modelling using Earth observation data * Understand popular Machine Learning methods and models to EO (Earth Observation) data * Recognize popular open-source foundation models available for fine tuning in EO data predictive modelling (transfer learning) |
| **METHODS** | Lectures, demos, code-alongs, application exercises, and structured discussion |
| **DURATION** | 11 hours for participants |

|  |  |  |  |
| --- | --- | --- | --- |
| # | SESSION | DURATION | LEARNING OBJECTIVES |
| 6.1 | Introduction to use cases in predictive modelling using EO/RS data | 3 hours | * Understand common use cases for predictive modelling using EO data analysis (including semantic segmentation of land use and crop type mapping, and crop yield prediction). * Identify the main challenges and opportunities associated with each common use case. * Identify the data requirements and sources for each common use case. * Be aware of additional use cases |
| 6.2 | Machine Learning methods & considerations **specific to EO data** | 6 hours | * Understand popular model architectures (algorithms) for ML using EO data and the various pros and cons of each (incl. CNNs, Random Forests, SVMs) * Understand variations on standard ML methods for training predictive models using EO data * Identify the key considerations and challenges associated with using EO data for predictive modelling (data preprocessing, feature selection, and model selection) * Understand considerations for assessing the performance of predictive models using EO data |
| 6.3 | Transfer learning (summary of popular open-source models available for fine tuning) | 2 hours | * Identify popular open-source models available for transfer learning in EO data analysis (including: VGG-16, ResNet, MobileNetV2 and Inception-v3) * Understand the main techniques and considerations associated with fine-tuning pre-trained models for specific EO data analysis tasks. |

**OVERVIEW OF EXERCISES**

|  |  |  |
| --- | --- | --- |
| **#** | **Exercise Name** | **Description** |
| Ex 6.1 | **Land-Use-Land-Cover** | Use of Google Earth Engine with Rwandan geodata to train a LULC classifier using various algorithms available in scikitlearn |
| Ex 6.2 | **Deep Learning for Crop Yield Estimation** | Use of Google Earth Engine with US geodata to train a CNN in pytorch to predict crop yield Crop yield prediction |
| Ex 6.3 | **Maize and nonmaize crop binary classification** | Use of Google Earth Engine with US geodata to train a random forest classifier to predict occurence of a specific crop (maize) |
| Ex 6.4 | **Field boundary delineation** | Use of RadiantMLHub with Rwanda geodata and keras to train a CNN to predict the boundaries of fields |
| Ex 6.5 | **Carbon stock prediction** | Use of Earthpy with Rwanda geodata and scikitlearn to train a linear regression model to predict carbon stock |

**6.1** **Introduction to use cases in predictive modelling using EO/RS data**

**6.1 Quiz questions**

|  |  |
| --- | --- |
| 1. Which of the following tasks can ML algorithms in remote sensing generally be grouped into? | 1. Classification and Regression 2. Clustering and Image Segmentation 3. Semantic Segmentation and Instance Segmentation 4. All of the above |
| 2. What tasks can ML classifiers be used for in remote sensing applications? | 1. Map land-cover and land-use 2. Map agricultural management systems 3. Detect event-like changes 4. All of the above |
| 3. Which ML algorithms are predominantly used in EO applications? | 1. Support Vector Machines (SVM), Random Forests (RF), and deep neural networks 2. K-Nearest Neighbors (KNN), Decision Trees, and Gradient Boosting 3. Naive Bayes, Logistic Regression, and AdaBoost 4. Linear Regression, Lasso Regression, and Ridge Regression |
| 4. What is the purpose of regression approaches in ML applications in remote sensing? | 1. Predict categorical labels for image pixels 2. Predict continuous output variables from EO data 3. Assign categorical labels to image pixels 4. Group pixels into homogeneous geospatial-objects |
| 5. What is the use of clustering in ML applications in remote sensing? | 1. Predict continuous output variables from EO data 2. Group pixels by feature similarity 3. Assign categorical “class” labels to single image pixels 4. Facilitate TPE in the exploration of the hyperparameter space by leveraging the collective knowledge of the swarm |
| 6. What is the significance of image segmentation in machine learning applications in remote sensing? | 1. Predict continuous output variables from EO data 2. Assign categorical “class” labels to single image pixels 3. Considered as a special way of clustering that accounts for spatial connectivity of neighbored pixels 4. Detect event-like changes |
| 7. What does an “area-adjusted” accuracy assessment report? | 1. Class-specific agreement considering the intersection under curve (IuC) 2. Average deviation of continuously predicted output values per reference area (e.g., average error of biomass in kg/ km²) 3. Class-specific map accuracies and error intervals, based on a statistically representative and independent validation sample |
| 8. Which metrics typically describe the performance of a regression? | 1. OA, UA, PA, F1 2. NDVI, EVI, LAI 3. MAE, RMSE, R² 4. IuC, AUC |

**Resources (articles / tutorials / videos):**

* Video/Workshop: [NASA ML4EO Workshop 2020 - YouTube](https://www.youtube.com/playlist?list=PL3QzFgBMGnbQRa8uHP0_C_P2Fl5GIBxmn)
* Video/Tutorial: [FAO Webinar Series: Earth observation data for agricultural statistics](https://www.fao.org/statistics/events/detail-events/en/c/1631683/)
* Video: [Hanna Meyer - Machine learning for earth observation - YouTube](https://www.youtube.com/watch?v=EyP04zLe9qo)
* Video/Online Courses: EO College:
  + Webpage <https://eo-college.org>
  + YouTube Channel <https://www.youtube.com/@EOCollege>
* Resource: Digital Earth Africa <https://maps.digitalearth.africa>
* Resource: [ESA Copernicus Open Access Hub](https://scihub.copernicus.eu/)
* Resource: Awesome Spectral Indices <https://awesome-ee-spectral-indices.readthedocs.io/en/latest/index.html>

**6.2** **Machine Learning methods & considerations specific to EO data**

**6.2 Quiz questions**

|  |  |
| --- | --- |
| 1. Why is the Random Forest (RF) algorithm heavily applied in remote sensing? | 1. It's the only algorithm available 2. Due to its performance, robustness, computational efficiency, and straightforward parametrization 3. Because it's easy to learn 4. None of the above |
| 2. What is the role of support vectors in Support Vector Machines (SVM)? | 1. They define the location of the hyperplane margin that separates different classes 2. They add noise to the data 3. They reduce the efficiency of the model 4. They decrease the computational time |
| 3. In Convolutional Neural Networks (CNN), what's the purpose of adding more convolutional layers? | 1. To recognize more complex objects 2. To reduce the complexity of the model 3. To decrease computational time 4. None of the above |
| 4. In evaluation of EO map data classification, what does the F1 score represent? | 1. The harmonic mean of precision and recall 2. The arithmetic mean of precision and recall 3. The geometric mean of precision and recall 4. The square root of the product of precision and recall |
| 5. Why is feature selection important in machine learning models? | 1. To reduce costs associated with data handling 2. To increase the model's performance by selecting the most informative features 3. To reduce the complexity of the model 4. All of the above |

**Resources (articles / tutorials / videos):**

Ref. 6.1

**6.3** **Transfer learning (summary of popular open-source models available for fine tuning)**

**6.3 Quiz questions**

|  |  |
| --- | --- |
| 1. What is one of the main reasons for using transfer learning in Earth Observation (EO) applications? | 1. It is easier to implement than training models from scratch. 2. It requires less computational resources compared to training a base model on a large image dataset. 3. It provides more accurate results than other types of learning. 4. It is the only method that works with EO data. |
| 2. Which of the following is an advantage of using ResNet over other backbones for semantic segmentation? | * 1. It can be implemented in PyTorch.   2. It involves skip connections to overcome issues with vanishing gradients.   3. It is the most complex implementation.   4. It is suitable for resource-limited devices. |
| 3. Which encoders is best-suited to resource-limited devices? | 1. VGG 2. ResNet 3. MobileNet 4. Inception |
| 4. What is one of the challenges when using pre-trained models with EO data? | 1. The models require large amounts of data for training. 2. The models need new inputs to come with the number and size of dimensions as the source data used for training. 3. The models do not perform well on large image datasets. 4. The models cannot handle multispectral data or time series. |
| 5. What is a common data augmentation routine in EO data? | 1. Reducing the size of the images. 2. Random rotations and horizontal/vertical flips of the images. 3. Changing the spectral information of the images. 4. Increasing the pixel size of the images. |

**Resources (articles / tutorials / videos):**

Ref. 6.1