Data Science for Geosciences

Project Session

Mathieu Fauvel

Monday 27 - Friday 31 January 2020

Project organization

- General objectives: Practice of machine learning tools for a real application
  - Statement of a given issue as a machine learning problem
  - Implementation of a machine learning model in Python (training & testing stage)
  - Evaluation and benchmarking of machine learning models
  - **Deliverable: Python notebook**
- Organization
  - 5 topics, 3 to 4 groups for each topic
  - Group of 2/3 trainees
  - Five project sessions
    1. Monday afternoon
    2. Wednesday afternoon
    3. Thursday afternonn
    4. Friday morning
    5. Friday afternoon

Topics

1. Prediction of ENSO (El Niño-Southern Oscillation)
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. Prediction of pollutant concentrations
6. Land cover classification from multitemporal Sentinel-2 Data

All projects resources available `https://github.com/DataScience4Geoscience/Toulouse2020/tree/master/Projects`

Sessions

- Session 1
  - Understanding of the targeted problem and of the associated data
  - Exploratory analysis and visualisation of the dataset
  - Statement of the targeted problem as a machine learning issue
- Session 2
  - Design of a machine learning framework for the targeted problem
  - Selection of ML models and libraries
  - Definition of evaluation criteria
  - 5' pitch by each project member to other trainees
- Session 3
  - Design your machine learning pipeline
- Session 4
  - Project synthesis (Python notebook)
  - Preparation of a joint presentation between the two groups for each project
- Session 5
  - Presentation of each project
  - Synthesis: question/response session
  - Feedbacks on the content and organization of the course

Prediction of ENSO (El Niño-Southern Oscillation)

Supervisors	P. Tandeo & R. Fablet
Data	Multivariate time series
Problem type	Regression
Keywords	Forecasting time series

Recognition of ocean-atmosphere events in satellite SAR images

Supervisors	P. Tandeo & R. Fablet
Data	SAR
Problem type	Classification
Keywords	Convolutional Neural Network

Analysis and dimensionality reduction of Sea Surface Height

Supervisors	L. Drumetz
Data	Occiput data
Problem type	Regression
Keywords	PCA, Autoencoders

Automatic fish sound recognition

Supervisors	M. Dalla Mura
Data	Accoustic data
Problem type	Classification
Keywords	Feature extraction, classification

Prediction of pollutant concentrations

Supervisors	F. Chatelain
Data	Multivariate time series
Problem type	Regression & Classification
Keywords	Random Forest, Knn, Variable selection

Land cover classification from multitemporal Sentinel-2 Data

Supervisors	M. Fauvel
Data	Satellite image time series
Problem type	Classification
Keywords	Feature extraction, classification