EO data processing in QGIS with a python API

Andreas Rabe, Benjamin Jakimow, Fabian Thiel, Sebastian van der Linden*

**andreas.rabe@geo.hu-berlin.de*

http://www.hu-geomatics.de

FOSS

Free Open Source

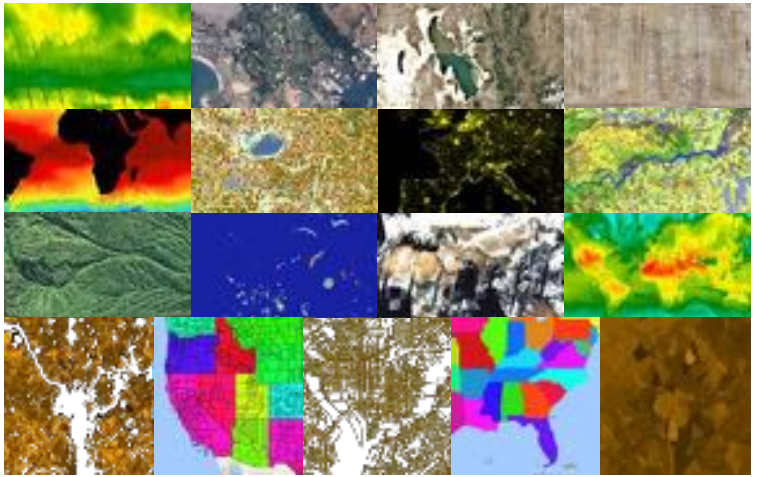
CSS

Closed Source

User

EO Data

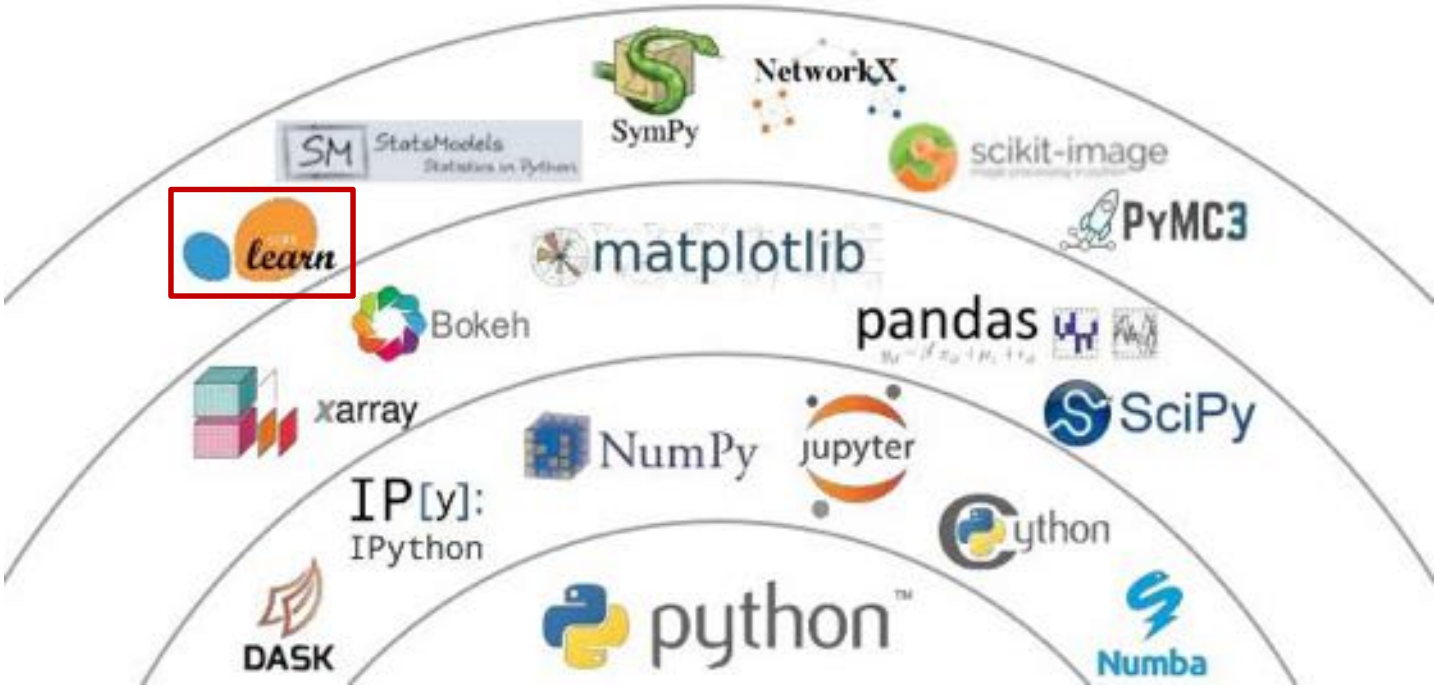
Climate, Imagery, Geophysical, Census



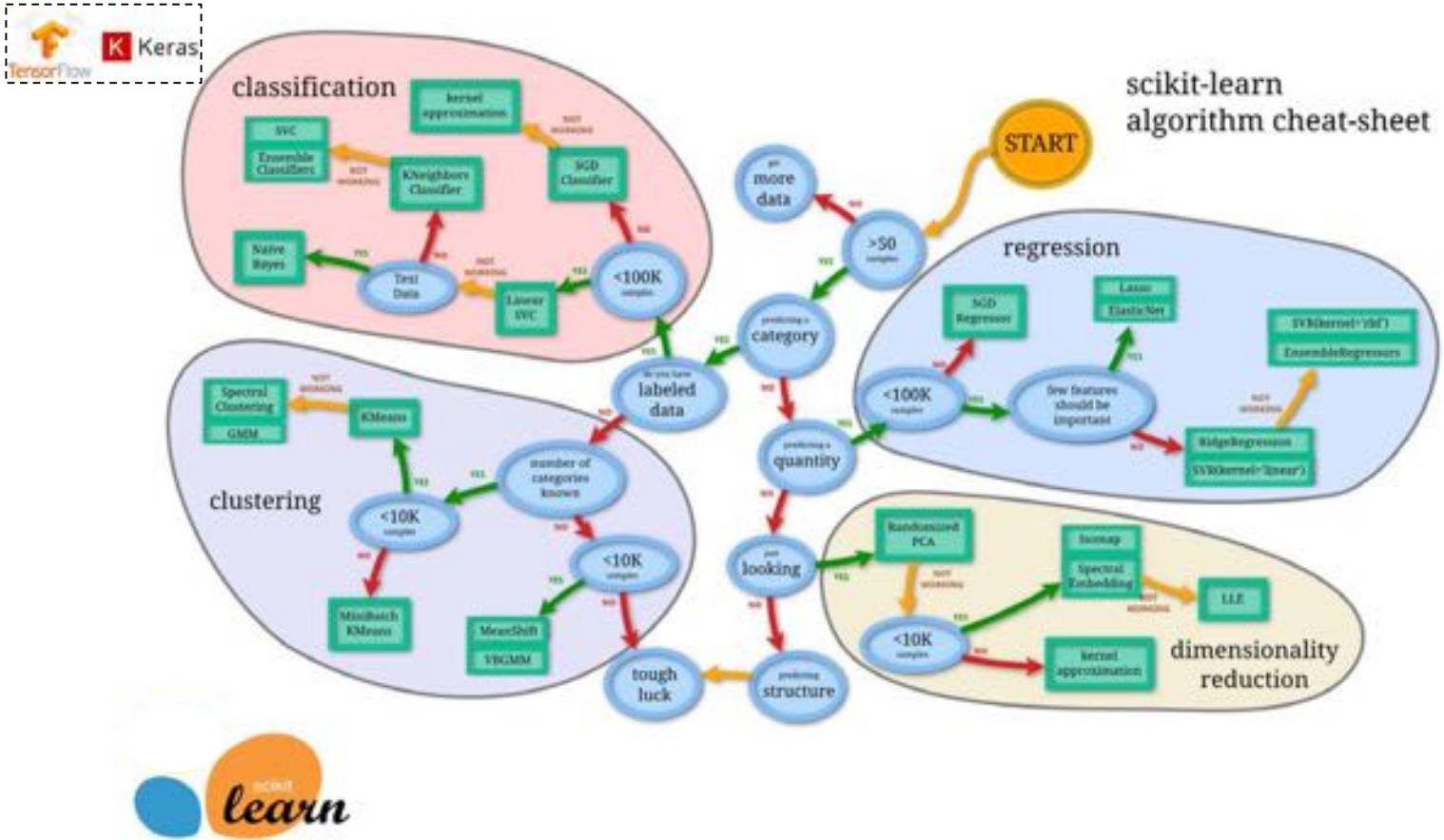
QGIS

GDAL Qt

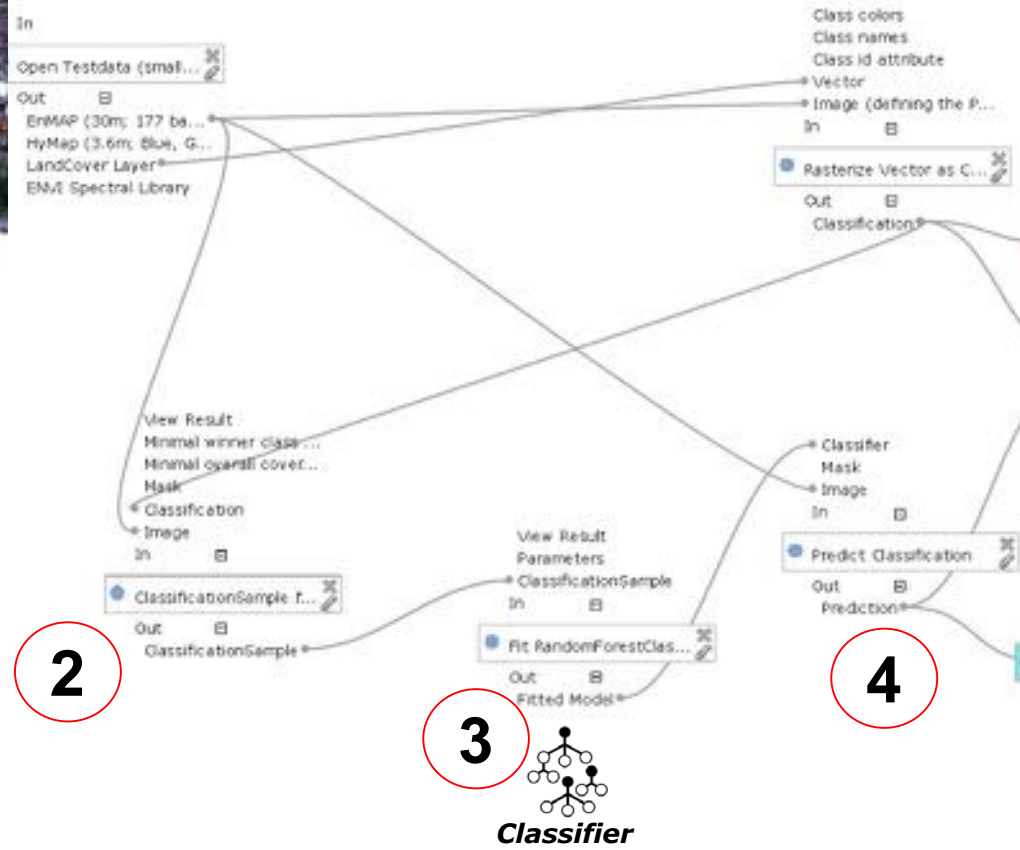
python



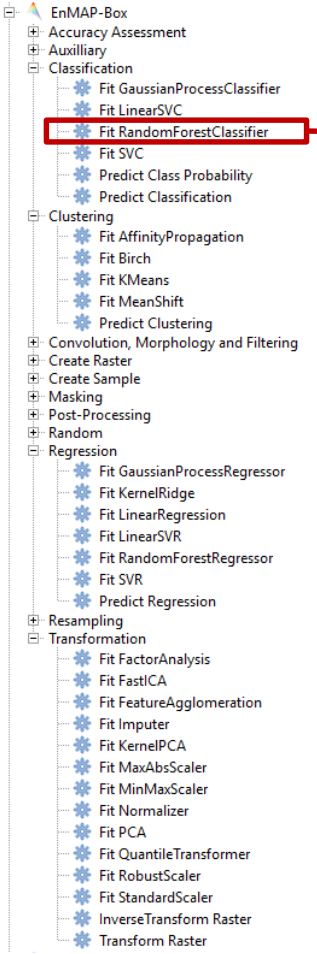
Scikit-Learn – Machine Learning in Python

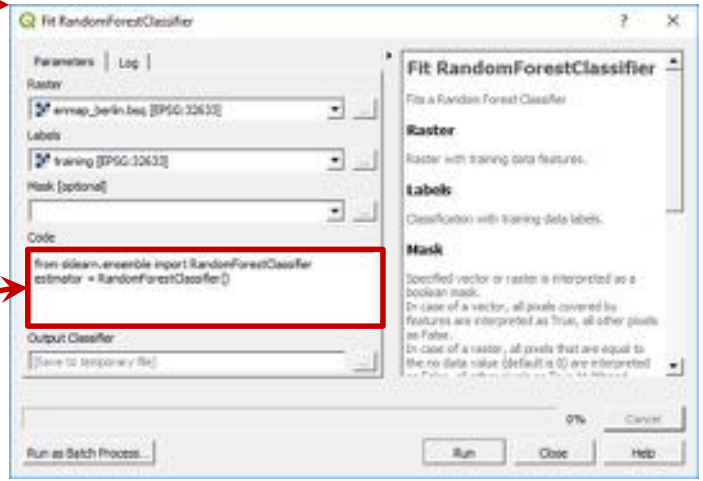


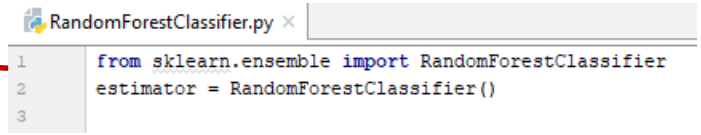
A typical Image Classification Workflow



Scikit-Learn Estimators in QGIS

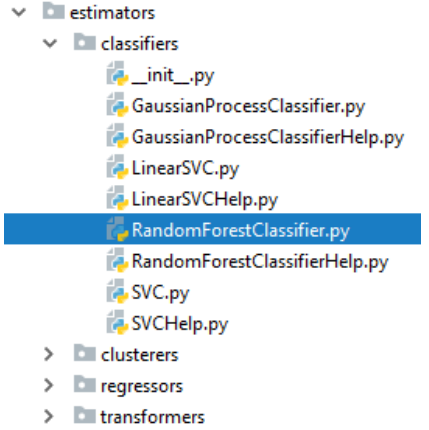


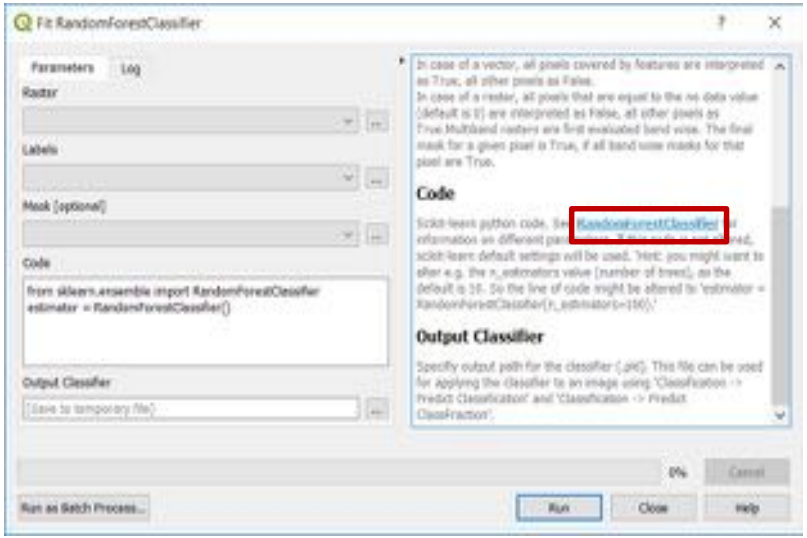




QGIS Processing Framework

scikit learn





3.2.4.3.1. sklearn.ensemble.RandomForestClassifier

class sklearn.ensemble. **RandomForestClassifier** (*n_estimators='warn', criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None*) [source]

A random forest classifier.

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True (default).

Read more in the User Guide.

Parameters:

n_estimators : *integer, optional (default=10)*

The number of trees in the forest.

Changed in version 0.20: The default value of `n_estimators` will change from 10 in version 0.20 to 100 in version 0.22.

criterion : *string, optional (default="gini")*

The function to measure the quality of a split. Supported criteria are "gini" for the Gini impurity and "entropy" for the information gain. Note: this parameter is tree-specific.

max_depth : *integer or None, optional (default=None)*

The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.

min_samples_split : *int, float, optional (default=2)*

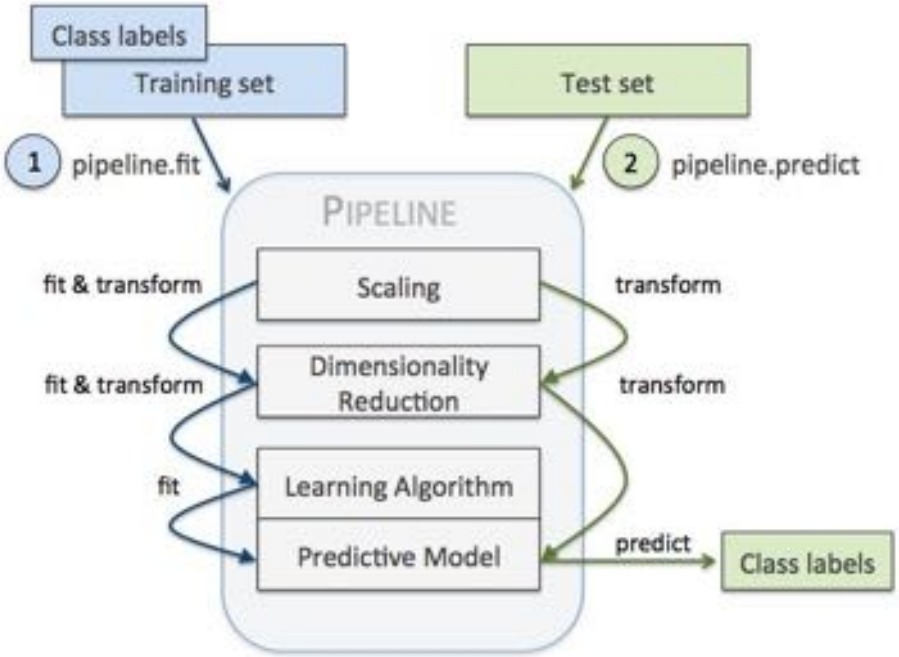
The minimum number of samples required to split an internal node:

- If int, then consider min_samples_split as the minimum number.
- If float, then min_samples_split is a fraction and ceil(min_samples_split * n_samples) are the minimum number of samples for each split.

Changed in version 0.18: Added float values for fractions.

min_samples_leaf : *int, float, optional (default=1)*

Pipeline Estimators



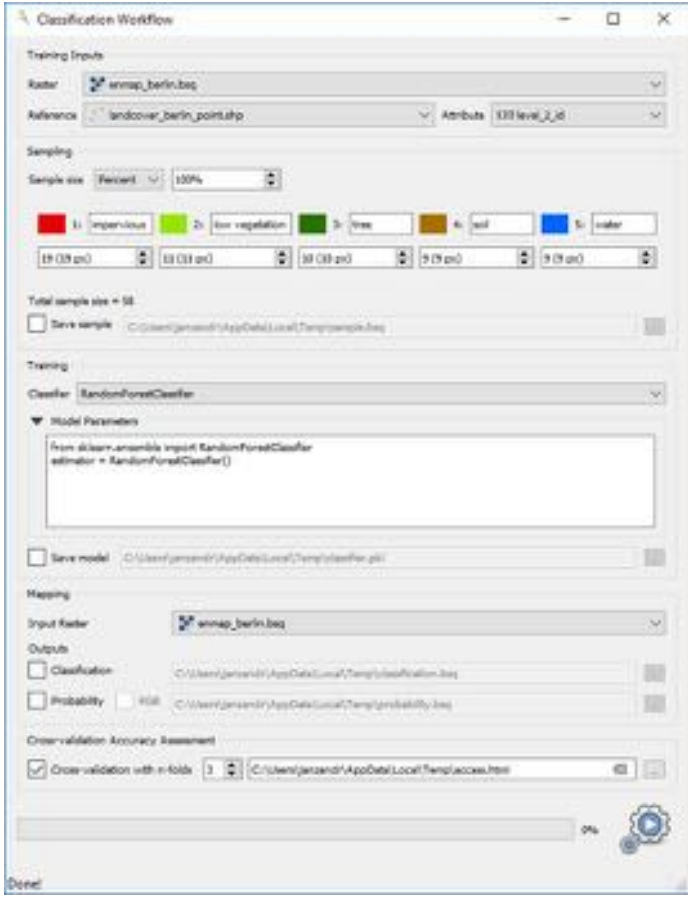
QGIS Processing Framework

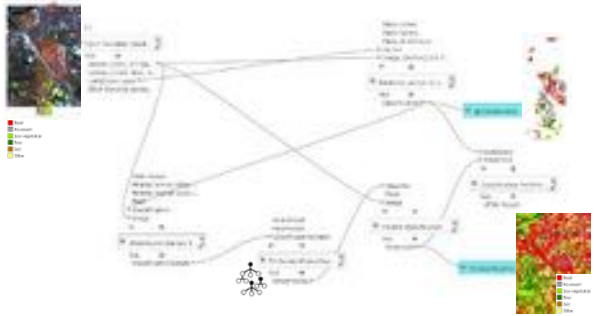
scikit learn

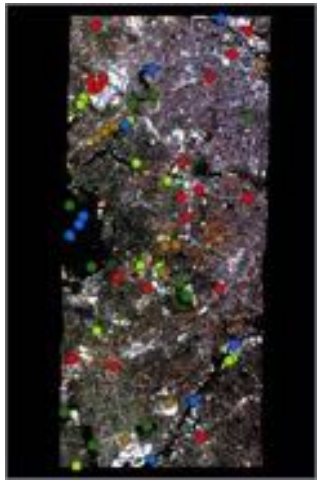
SVC.py

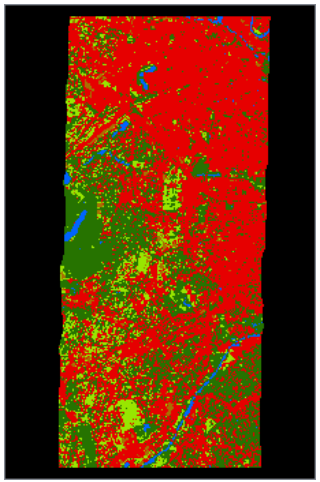
```
from sklearn.pipeline import make_pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC

svc = SVC(probability=False)

param_grid = {'kernel': ['rbf'],
              'gamma': [0.001, 0.01, 0.1, 1, 10, 100, 1000],
              'C': [0.001, 0.01, 0.1, 1, 10, 100, 1000]}

tunedSVC = GridSearchCV(cv=3, estimator=svc, scoring='f1_macro',
                        param_grid=param_grid)

estimator = make_pipeline(StandardScaler(), tunedSVC)
```

Interactive Applications





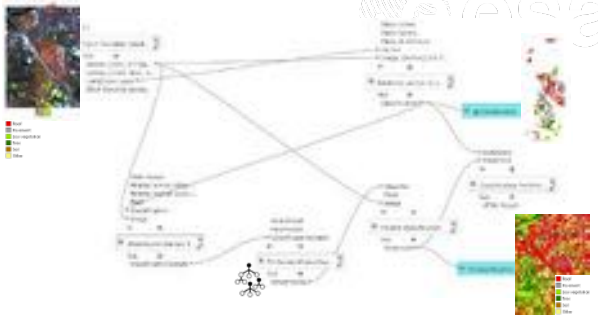




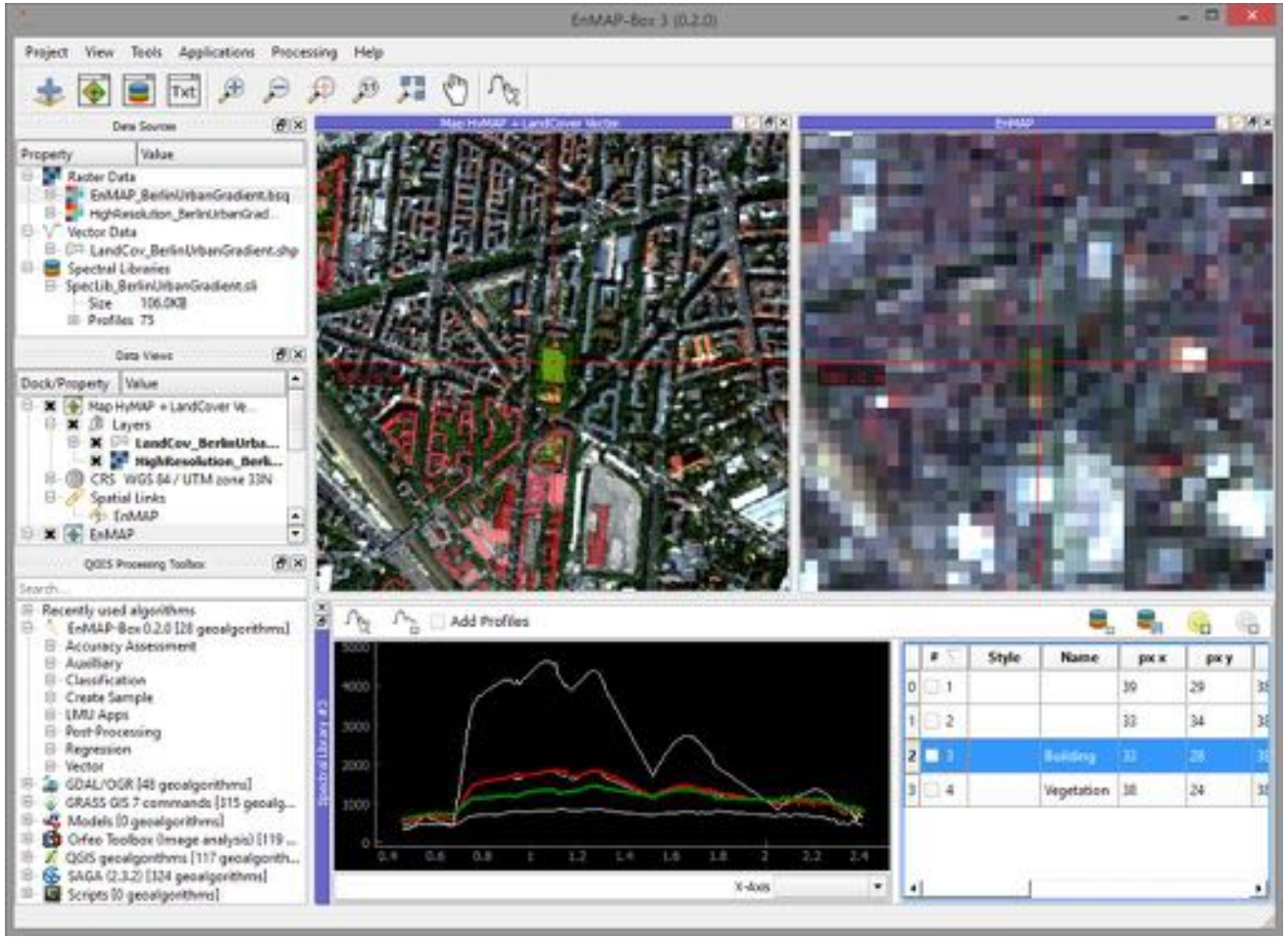
Confusion Matrix

	Reference Class					
	(1)	(2)	(3)	(4)	(5)	Sum
(1) impervious	15	2	0	1	0	18
(2) low vegetation	1	9	0	0	0	10
(3) tree	0	0	10	0	0	10
(4) soil	3	0	0	8	0	11
(5) water	0	0	0	0	9	9
Sum	19	11	10	9	9	58

Scripting the Workflow

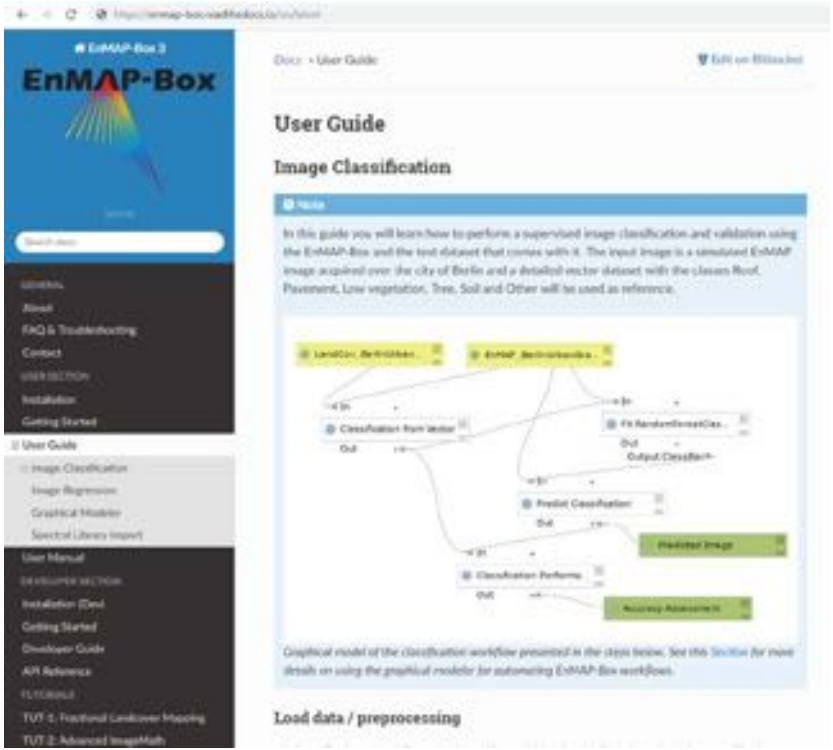


```
from hubflow.core import *
from sklearn.ensemble import RandomForestClassifier
import enmapboxtestdata

enmap = Raster(filename=enmapboxtestdata.enmap)

vectorClassification = VectorClassification(filename=enmapboxtestdata.landcover,
                                            classAttribute='Level_2_ID')

classification = Classification.fromClassification(classification=vectorClassification,
                                                   grid=enmap.grid(),
                                                   filename='/vsimem/classification.bsq')

sample = ClassificationSample(raster=enmap, classification=classification)

rfc = Classifier(sklEstimator=RandomForestClassifier())
rfc.fit(sample=sample)
rfc.predict(raster=enmap, filename='/vsimem/rfcClassification.bsq')
```

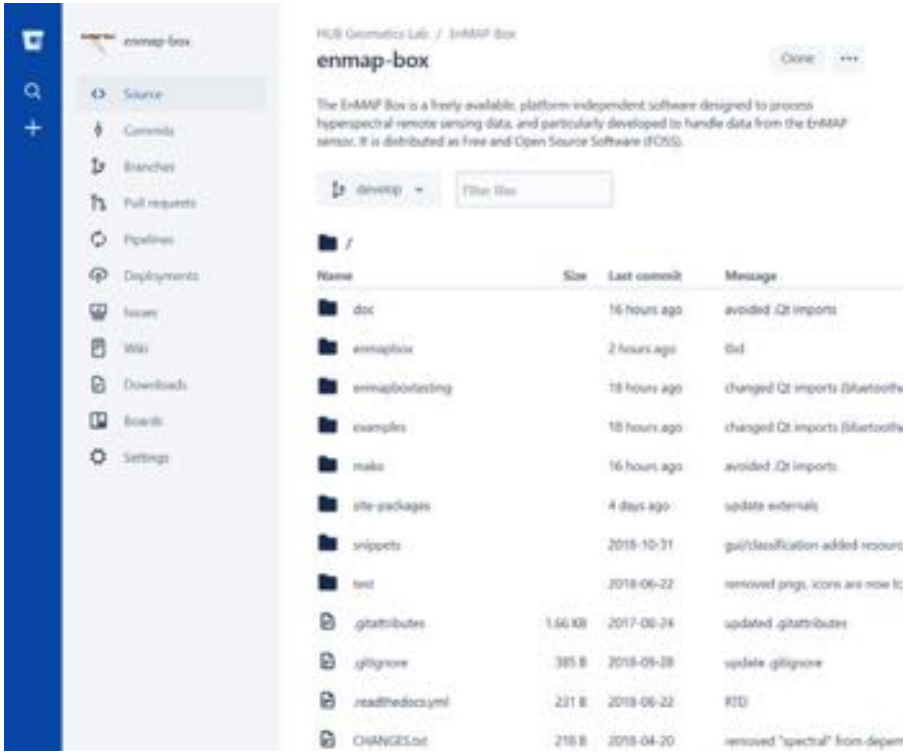

EnMAP-Box 3 (0.2.0)
Project View Tools Applications Processing Help
Data Sources
Property Value
Raster Data
EnMAP_BerlinUrbanGradient.bsq
HighResolution_BerlinUrbanGrad...
Vector Data
LandCov_BerlinUrbanGradient.shp
Spectral Libraries
SpecLib_BerlinUrbanGradient.sli
Size 106.0KB
Profiles 75
Data Views
Dock/Property Value
Map HyMAP + LandCover Ve...
Layers
LandCov_BerlinUrba...
HighResolution_Berli...
CRS WGS 84 / UTM zone 33N
Spatial Links
EnMAP
EnMAP
QGIS Processing Toolbox
Search...
Recently used algorithms
EnMAP-Box 0.2.0 [28 geoalgorithms]
Accuracy Assessment
Auxiliary
Classification
Create Sample
LMU Apps
Post-Processing
Regression
Vector
GDAL/OGR [48 geoalgorithms]
GRASS GIS 7 commands [315 geoalg...
Models [0 geoalgorithms]
Orfeo Toolbox (image analysis) [119 ...
QGIS geoalgorithms [117 geoalgorith...
SAGA (2.3.2) [124 geoalgorithms]
Scripts [0 geoalgorithms]
Map HyMAP + LandCover Vector
EnMAP
Add Profiles
Spectral Library #3
5000
4000
3000
2000
1000
0
0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4
X-Axis
Style Name px x px y
0 1 39 29
1 2 33 34
2 3 Building 33 28
3 4 Vegetation 38 24

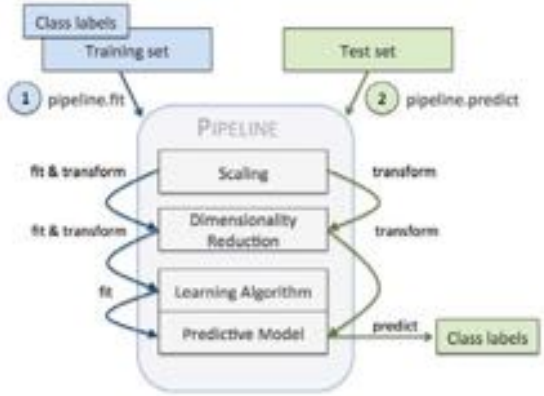

About EnMAP-Box | About
EnMAP-Box 3
Version 3.3.20181109T1242.develop
Website, Source Code and more
https://bitbucket.org/hu-geomatics/enmap-box
Licenced under the GNU General Public Licence
http://www.gnu.org/licenses/
Environmental Mapping and Analysis Program (EnMAP)
http://www.enmap.org
The EnMAP-Box is developed at Humboldt-Universität zu Berlin under contract by the Helmholtz Centre Potsdam GFZ and is part of the EnMAP Core Science Team activities. It is funded by the German Aerospace Centre (DLR) - Project Management Agency, granted by the Federal Ministry of Economic Affairs and Energy (BMWi; grant no. 50EE1529).
GFZ
Helmholtz-Zentrum
POTSDAM
DLR

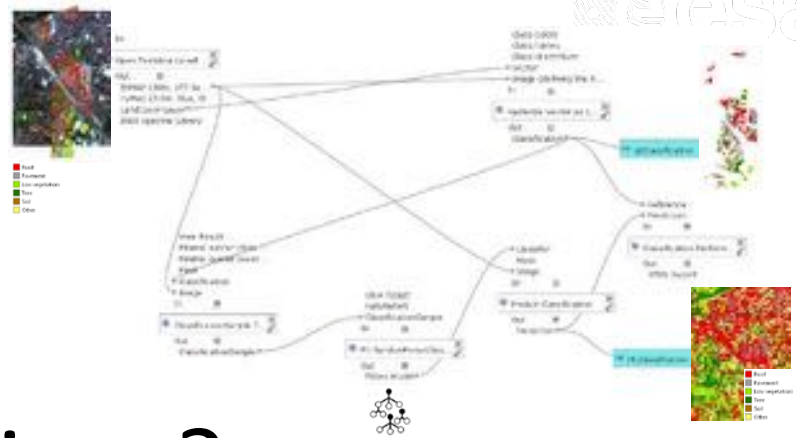
Read the Docs
https://enmap-box.readthedocs.io



Bitbucket
https://bitbucket.org







Any questions?