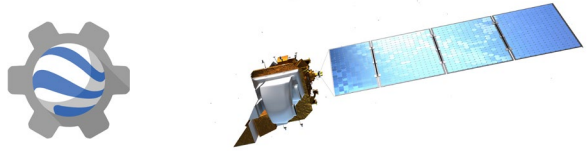
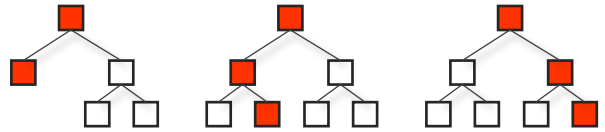
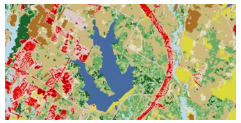
Supervised Land-Use/Land-Cover (LULC) Classification

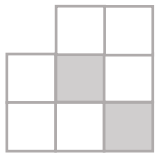
Prepared by Dr. Brad G. Peter, Department of Geography, University of Alabama

- Access multispectral Landsat 8 imagery and the USGS National Land Cover Database (NLCD) through Google Earth Engine
- Build a training dataset for classifying different LULC types
- Use a **machine learning algorithm** to construct a spatially continuous LULC model
- Evaluate a confusion matrix to determine model overall accuracy
- Re-parameterize the algorithm to maximize predictive accuracy

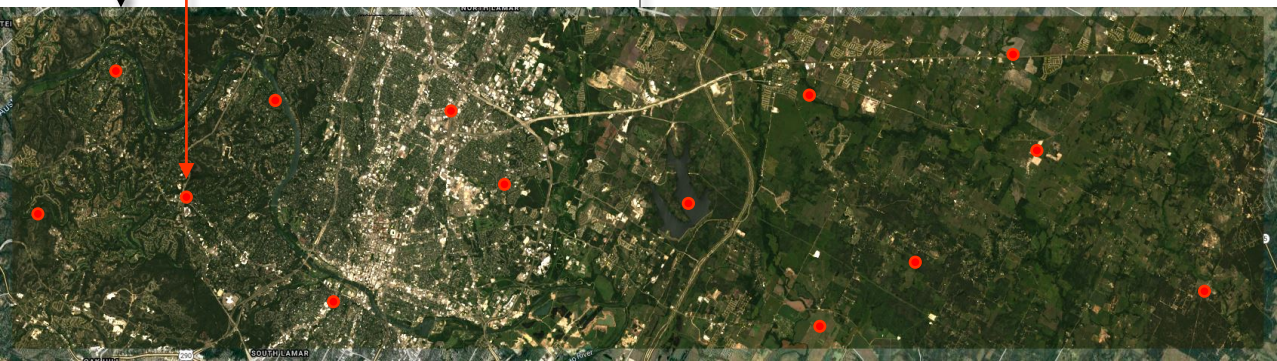


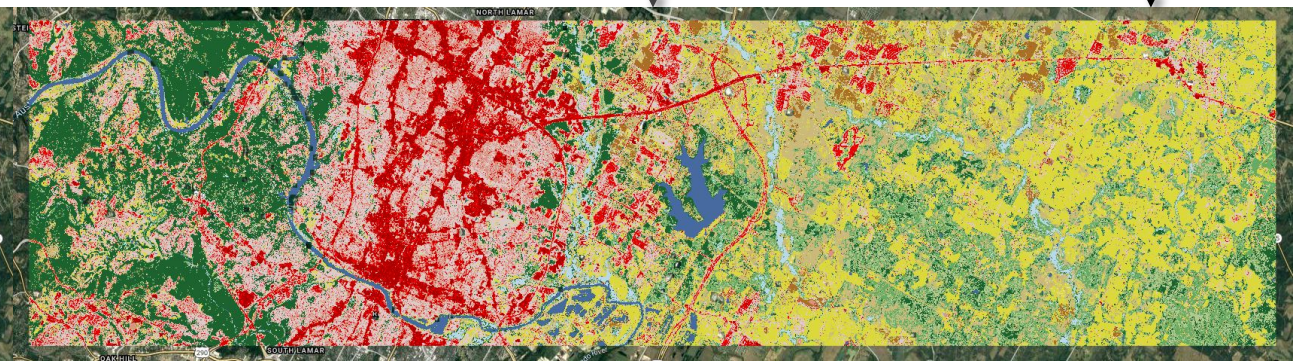






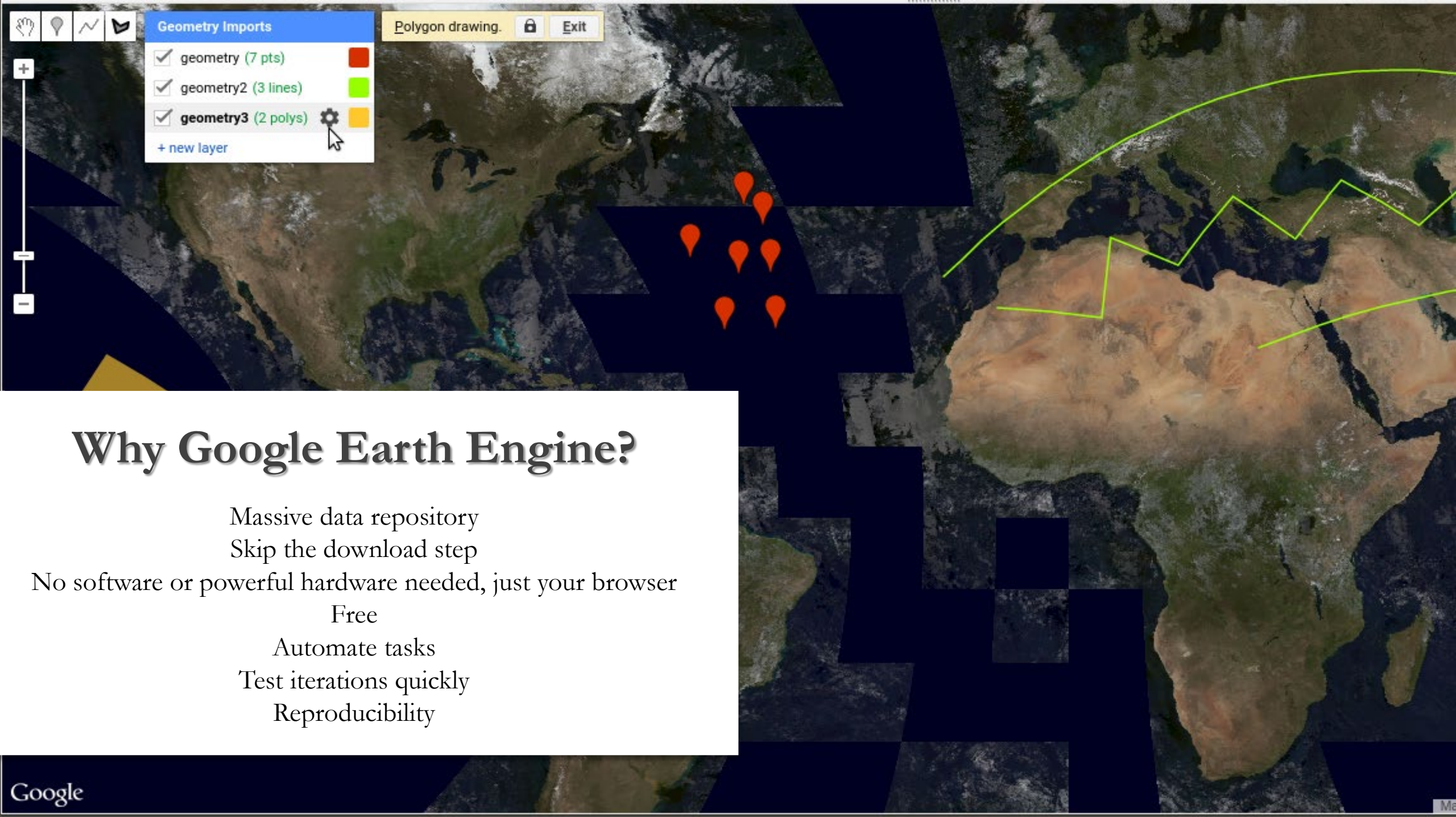
Random forest algorithm





Description and objectives

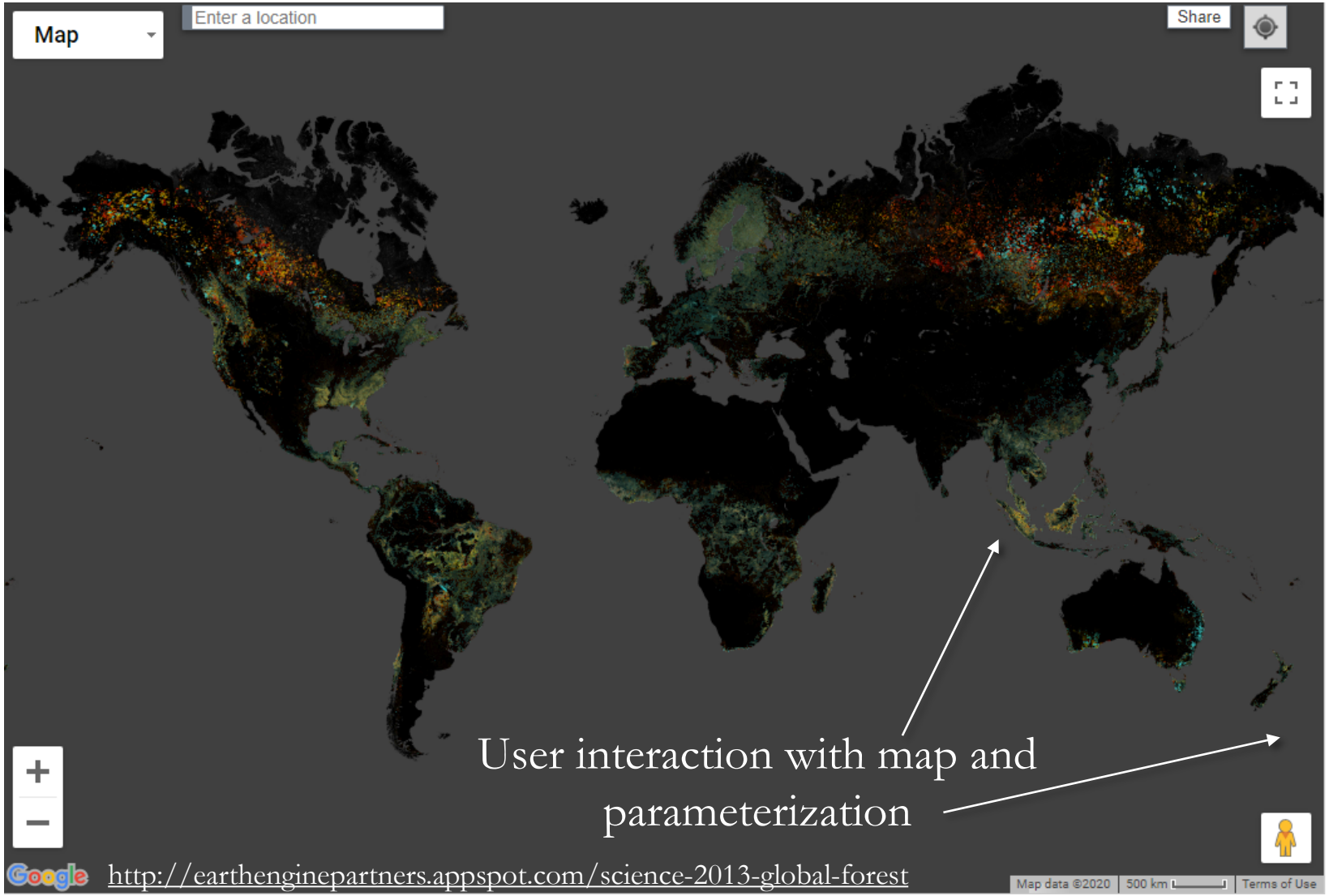
- Learn to create and troubleshoot supervised land-use/land-cover (LULC) classifications using Google Earth Engine (GEE)
- You will use multispectral optical imagery from the Landsat 8 satellite to train a supervised classification model and compare it to the USGS National Land Cover Database
- Classification accuracy will be calculated using the results of a confusion matrix and a validation layer.
- **You will need a GEE account to participate!**



Why Google Earth Engine?

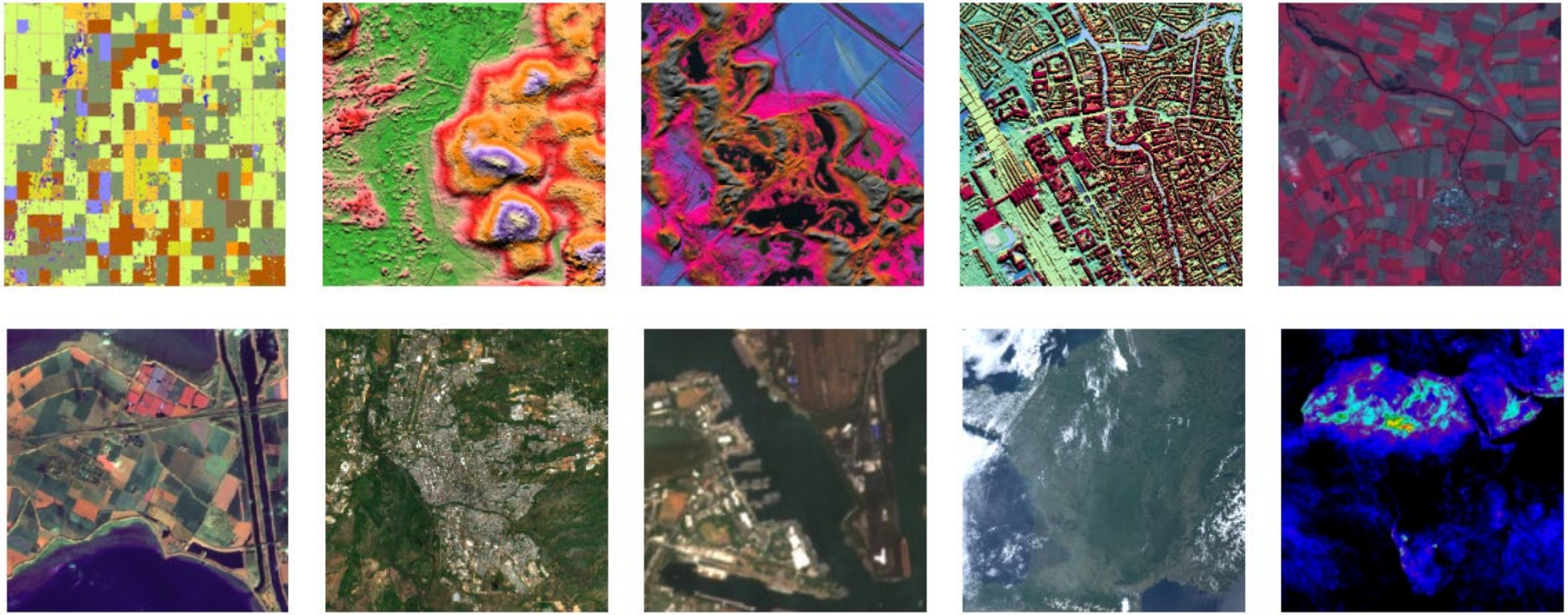
Massive data repository
Skip the download step
No software or powerful hardware needed, just your browser
Free
Automate tasks
Test iterations quickly
Reproducibility

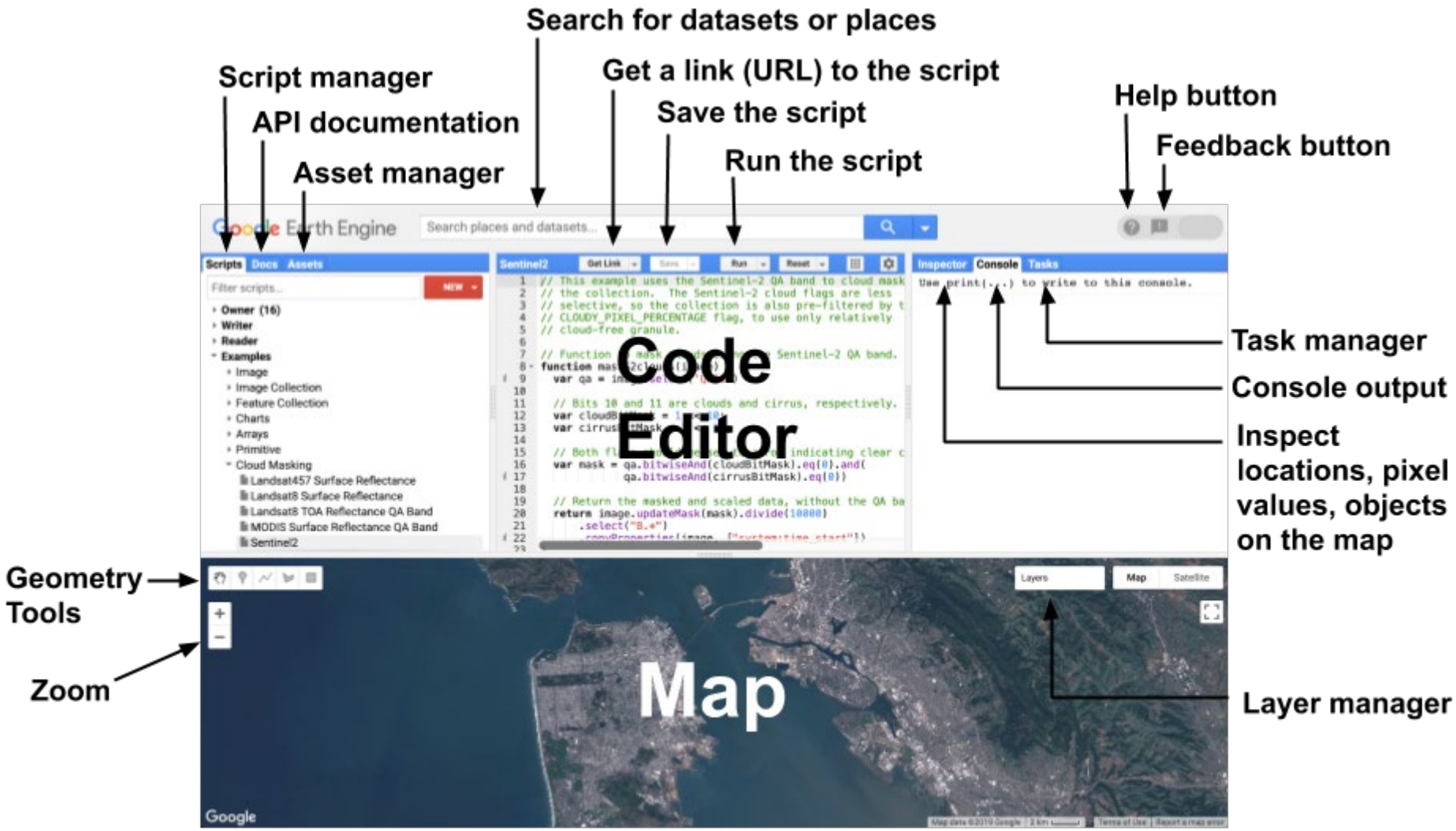
Monitoring global forest change



Where do I get the data?

- No need to download! Access on the cloud:
- https://developers.google.com/earth-engine/datasets/catalog





https://developers.google.com/earth-engine/playground

Why classify?

What else?

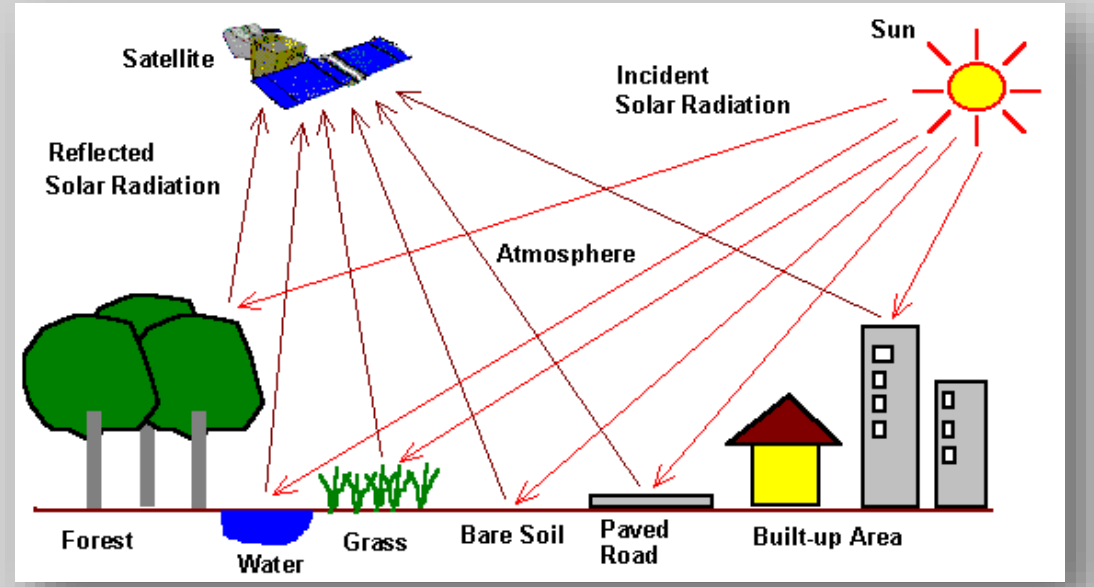
- LULC data are needed for many applications, such as wildlife conservation, natural resource management, urban planning, and monitoring agriculture

- One of the benefits of remote sensing data for this purpose is that satellite images offer **spatially continuous** data at regular revisit rates

What does this mean?

Surface reflectance refresher

- Multispectral optical imagery contains a wealth of information that can be used to distinguish features on the earth's surface
- Solar radiation travels to the earth and wavelengths of light are absorbed and reflected differently by different materials/objects
- The reflected solar radiation is what is captured by the satellite
- These data can then be used to classify land-use/land-cover based on the spectral signatures of the objects.

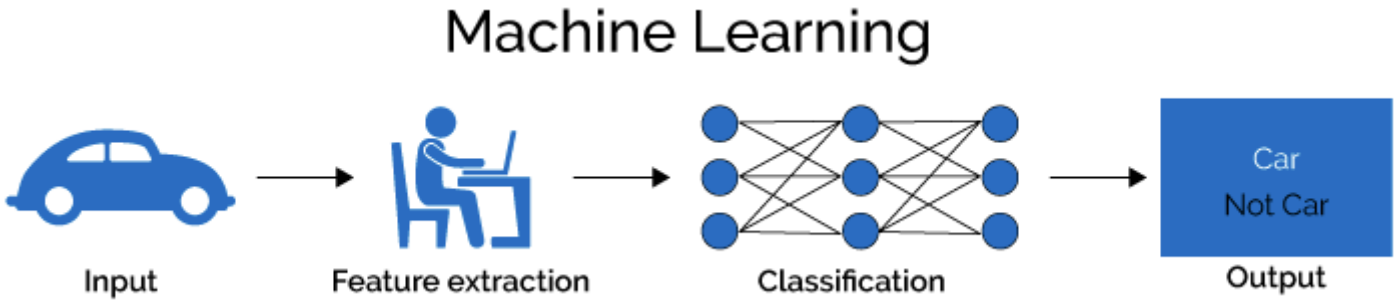


Machine learning

- **Oxford definition:**
- "The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data."

- Have you heard of unsupervised/supervised classification models?

https://lawtomated.com/a-i-technical-machine-vs-deep-learning/

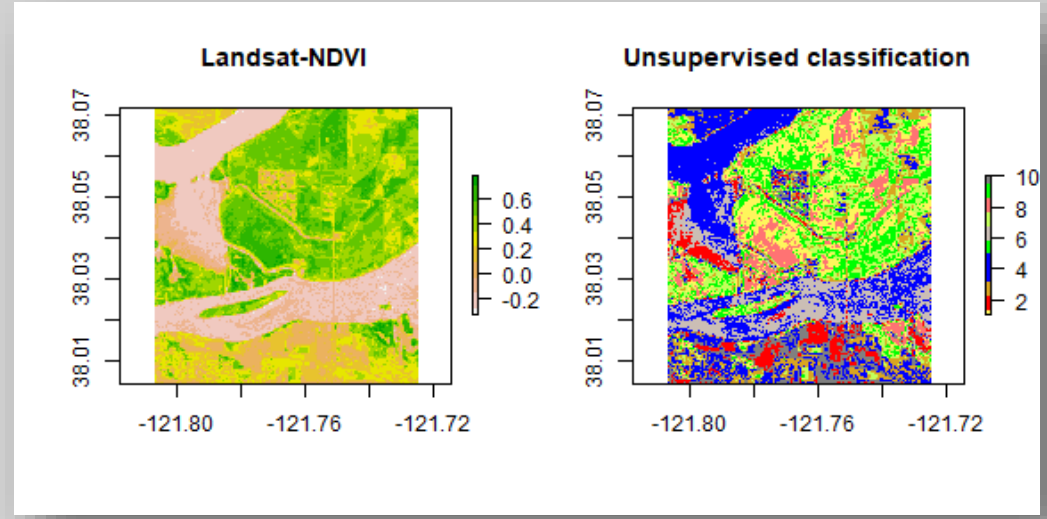


Machine learning algorithms in GISci

Supervised	Unsupervised
Linear regression	K-means
Decision tree	Principal Component Analysis (PCA)
Random forest	
CART	
Support Vector Machine (SVM)	

What is unsupervised classification?

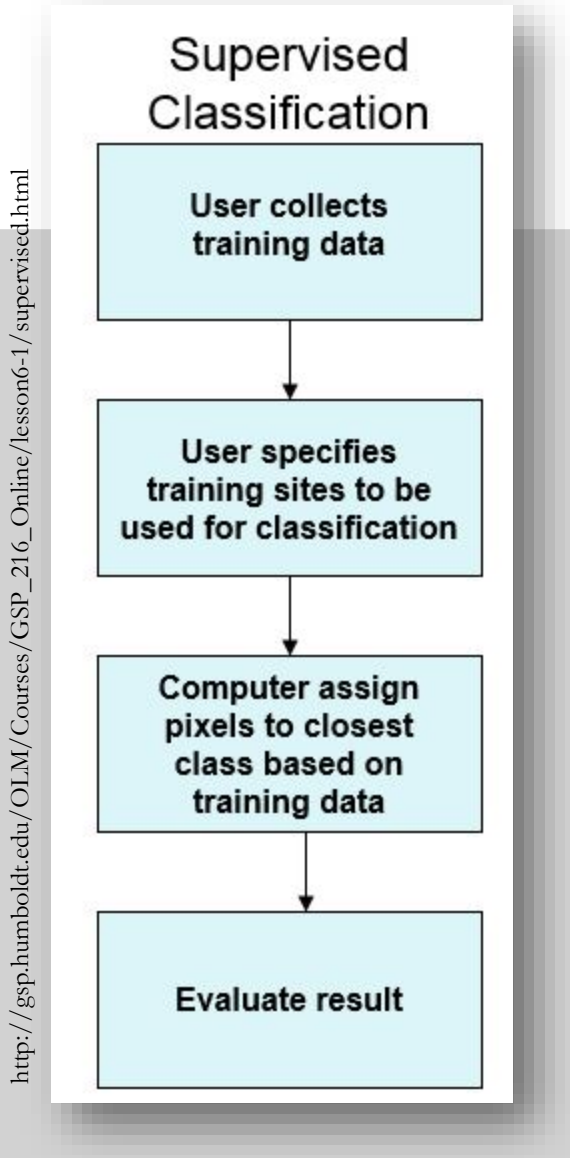
- Algorithms use untrained data to identify patterns and cluster features into groups – more than one variable can be used for clustering
- Output classifications are **unknown**
- **Popular method:** K-means clustering

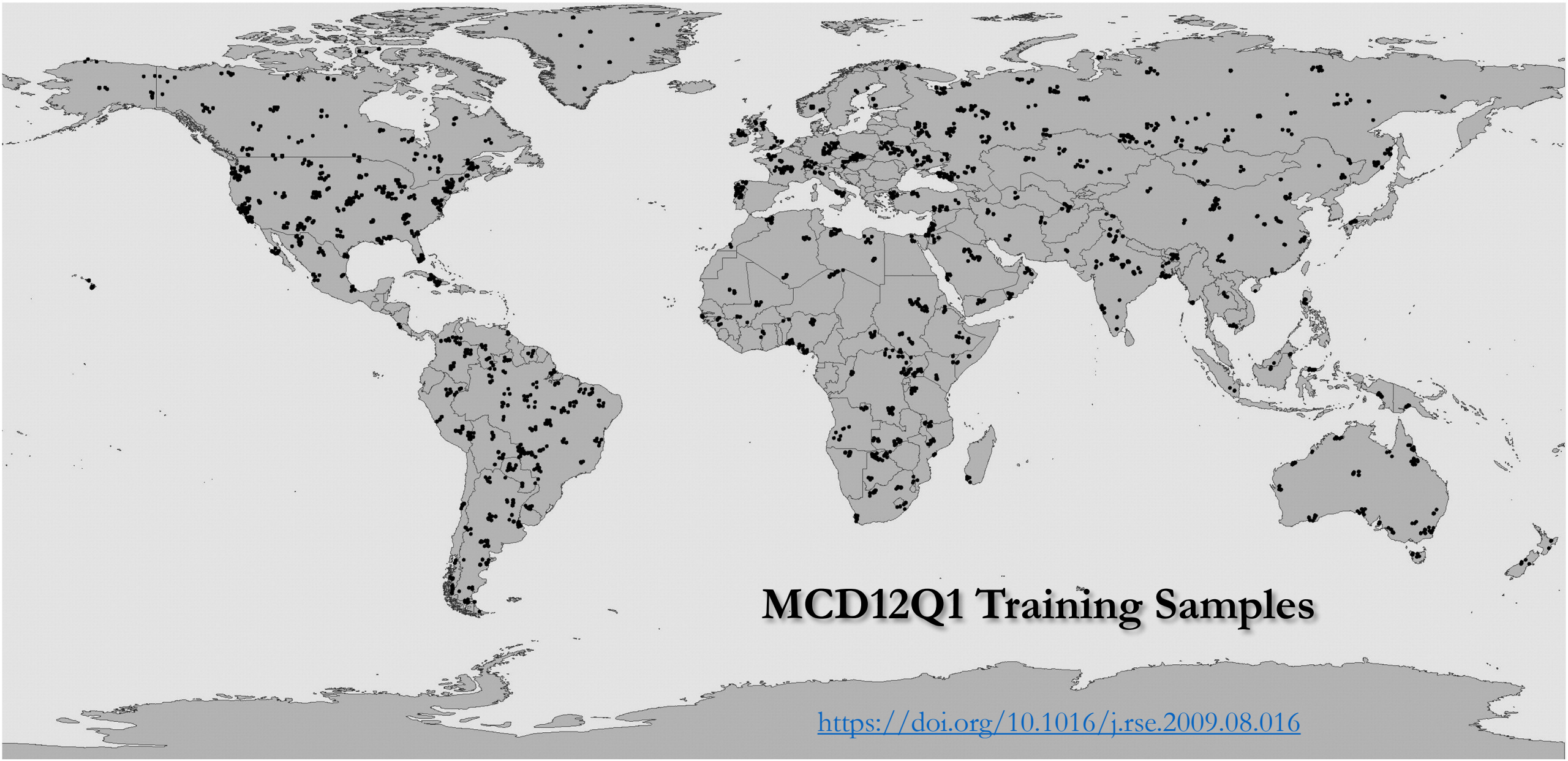


https://rspatial.org/raster/rs/4-unsupclassification.html

What is supervised classification?

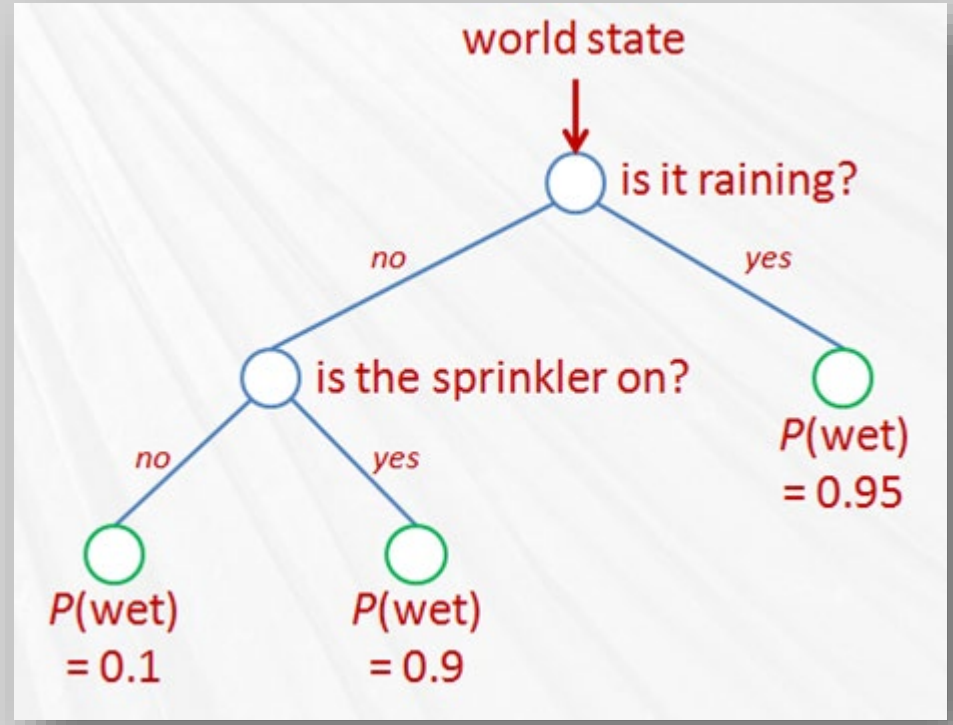
- Supervised/unsupervised classification methods
- Supervised classification is a technique in which samples with **known values** are used to train a classification algorithm
- We'll be accessing the spectral signatures of points across the USGS NLCD dataset and use the reflectance data to train a model that predicts spatially continuous LULC
- Validated with data that were not used in model training




MCD12Q1 Training Samples
https://doi.org/10.1016/j.rse.2009.08.016

Classification and regression tree (CART)

- **What is it?**
  - Series of decisions based on regression results from training data

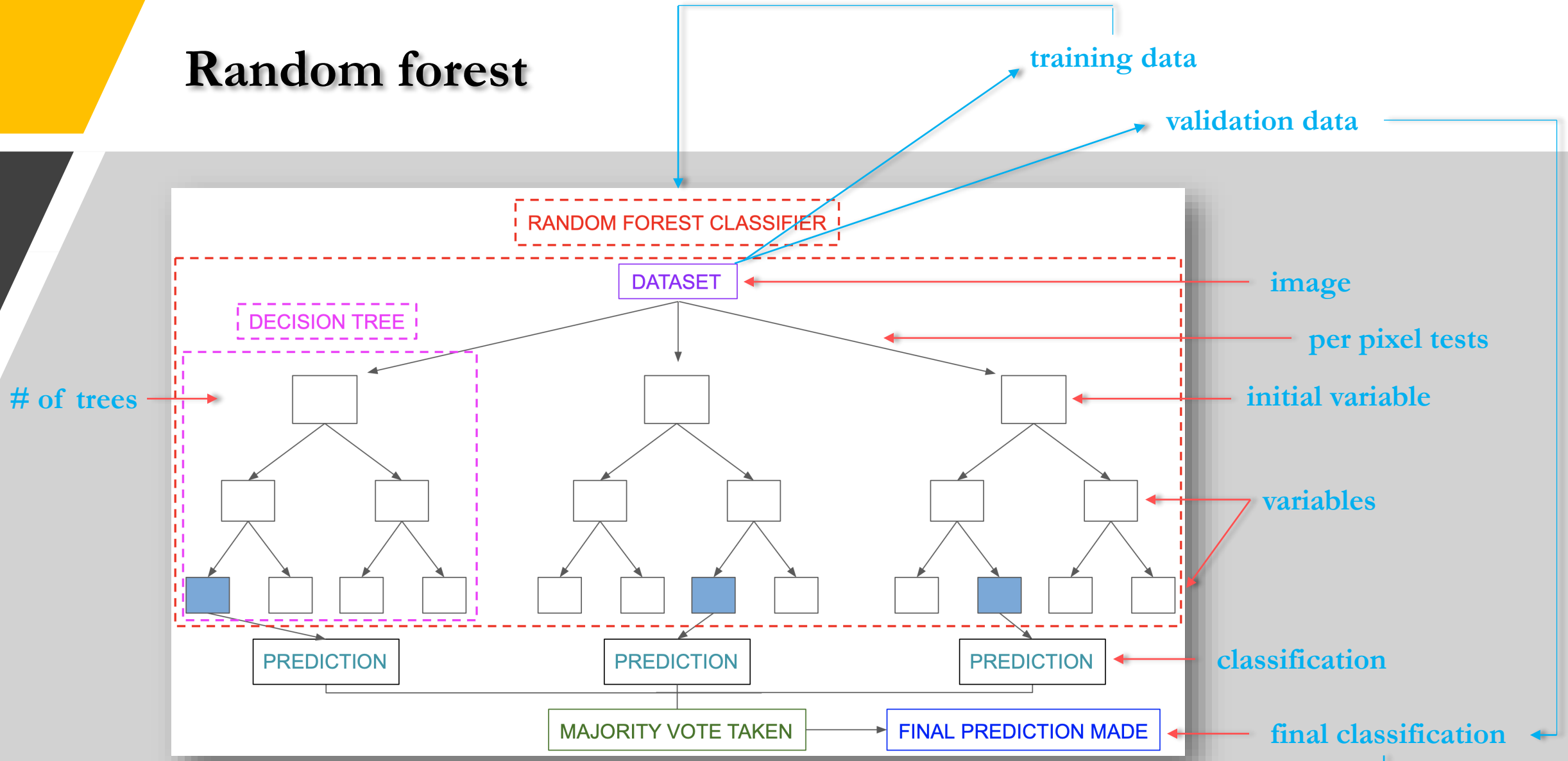


https://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:random_forests

If you want to test:

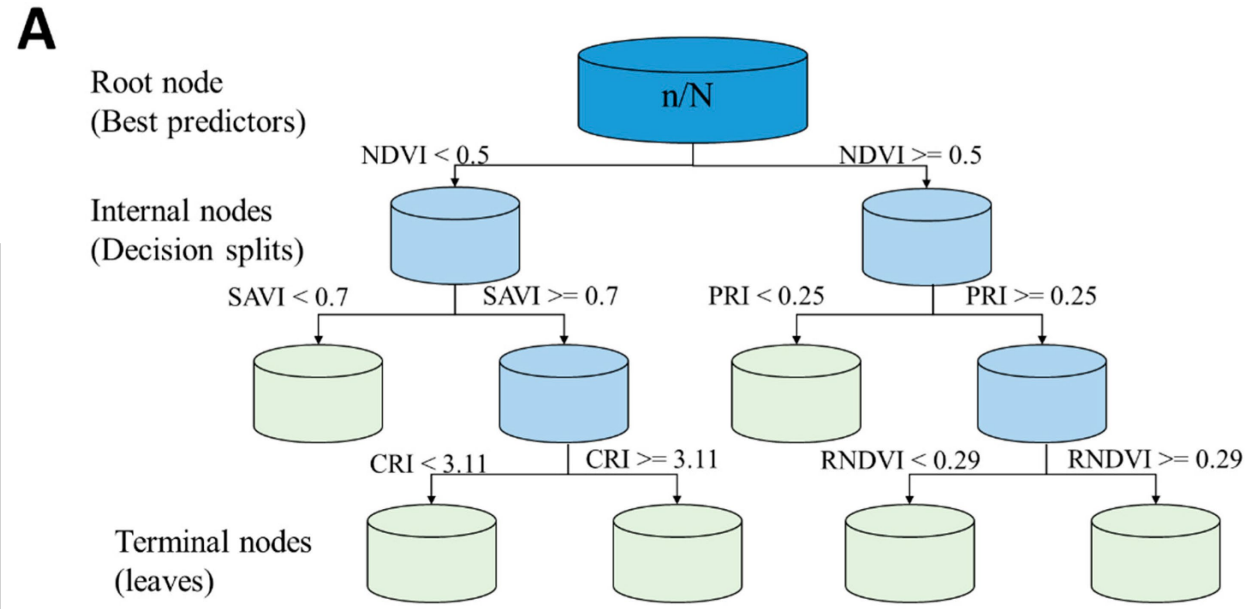
```
var cart = ee.Classifier.smileCart().train({
  features:training,
  classProperty:'landcover',
  inputProperties: bands
})
print(cart.explain())
var cartMap = masked.classify(cart)
Map.addLayer(cartMap,{min:0,max:19,palette:classColors},'CART',false)
```

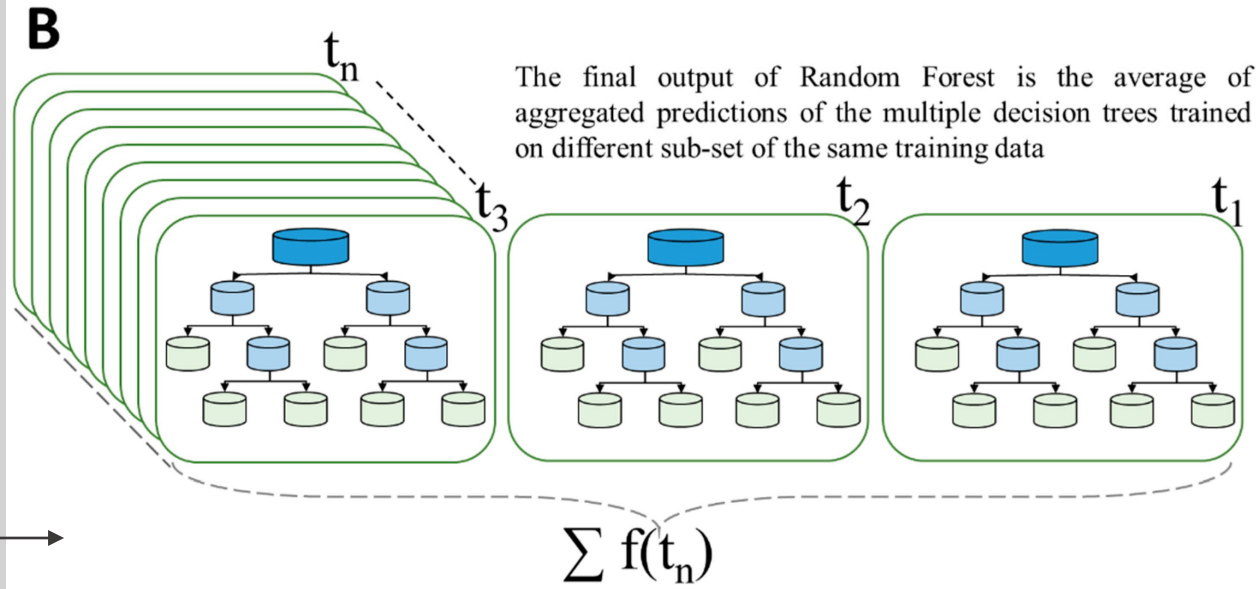
Random forest



https://medium.com/analytics-vidhya/machine-learning-decision-trees-and-random-forest-classifiers-81422887a544

Random forest with remote sensing indices

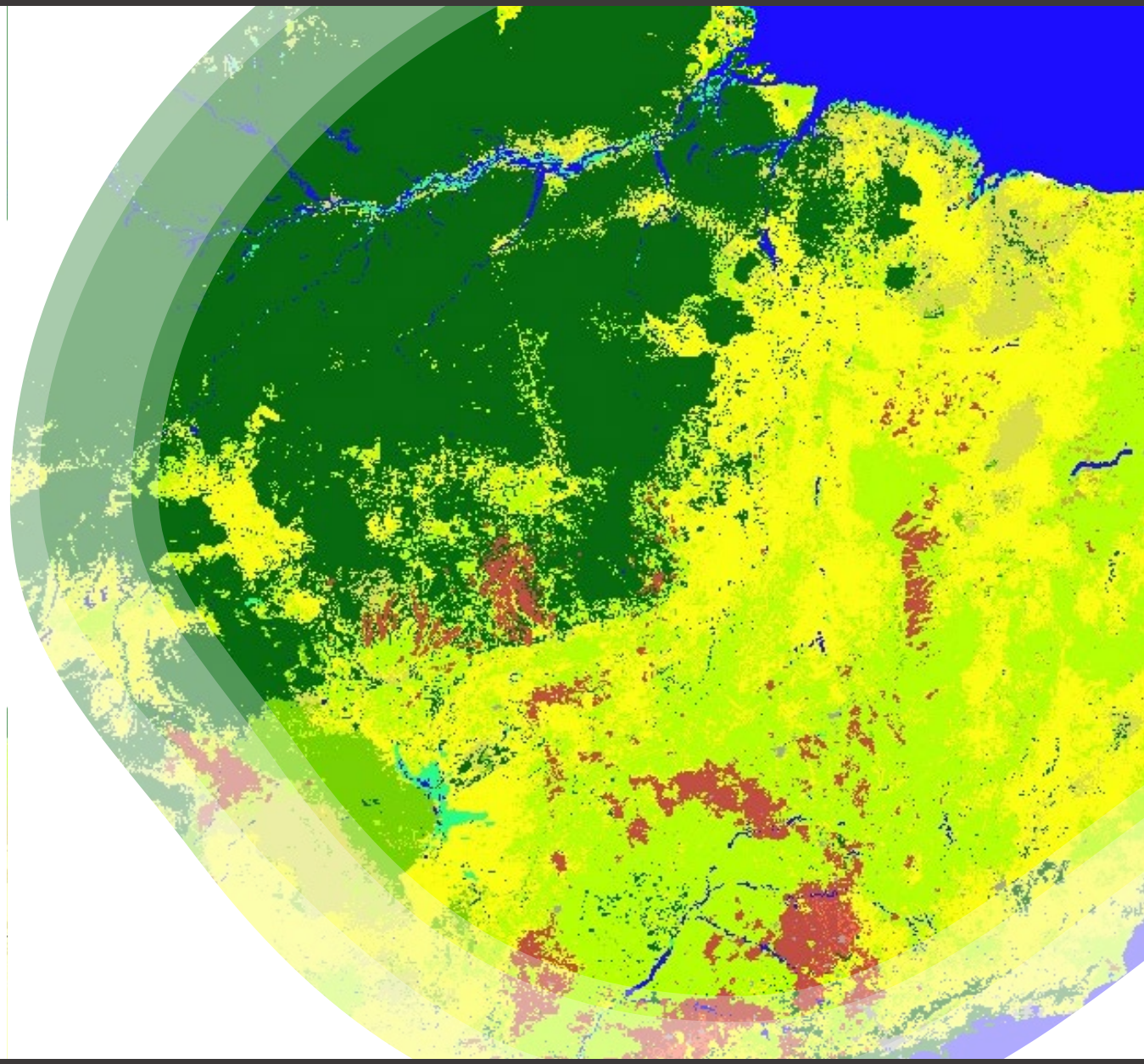




https://www.mdpi.com/2072-4292/11/8/920/htm

Google Earth Engine Challenge

- https://github.com/ML4SCIHackathon/ML4SCI/tree/main/GoogleEarthEngineChallenge
- DUE 10/30/2020 by 6:00pm CST



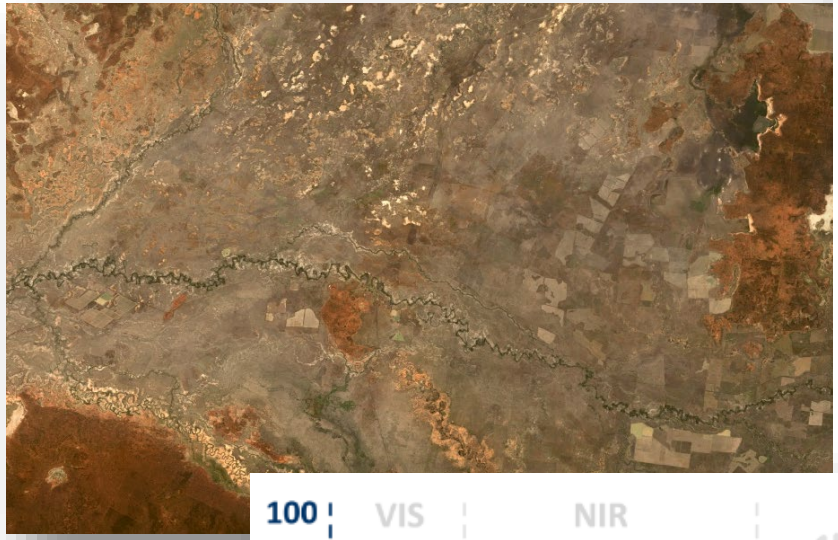
NLCD 2016 Landcover

NLCD 2016 Land Cover for the conterminous United States represented as 16 land cover classes.

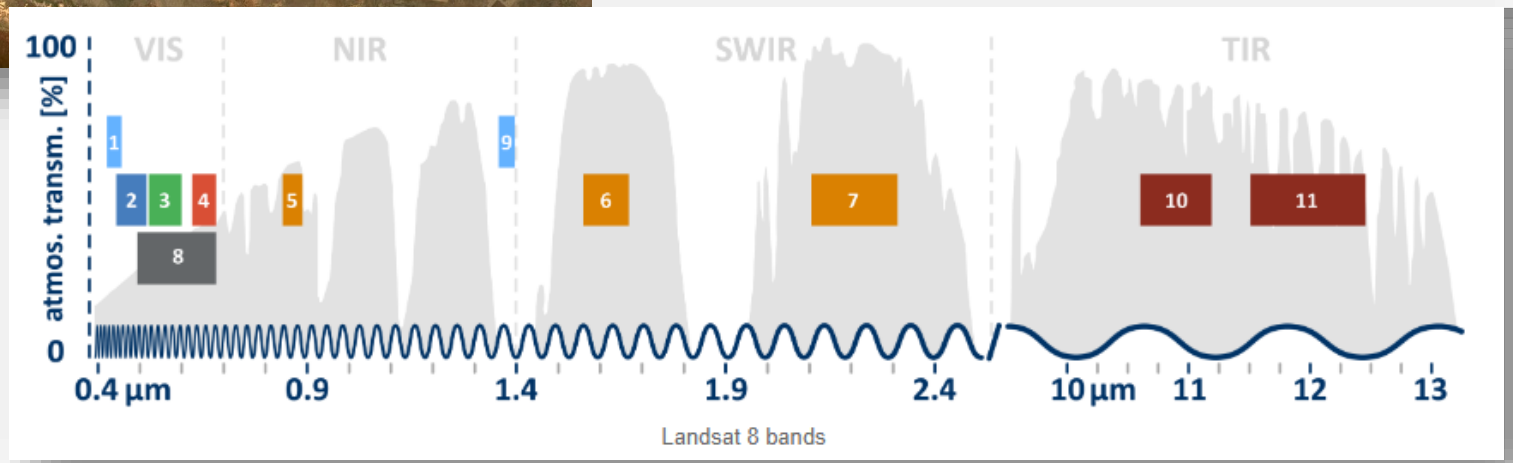
Key to Land Cover Types

- Open Water
- Perennial Ice and Snow
- Developed, Open Space
- Developed, Low Intensity
- Developed, Medium Intensity
- Developed, High Intensity
- Barren Land
- Deciduous Forest
- Evergreen Forest
- Mixed Forest
- Dwarf/Scrub
- Shrub/Scrub
- Grassland/Herbaceous
- Sedge/Herbaceous
- Moss
- Pasture/Hay
- Cultivated Crops
- Woody Wetlands
- Emergent Herbaceous Wetlands

Landsat 8 Surface Reflectance



BAND	SPECTRAL	WAVELEN. [µm]	GEOM. [m]	SENSOR
1	aerosols	0.435 – 0.451	30	OLI
2	blue	0.452 – 0.512	30	OLI
3	green	0.533 – 0.590	30	OLI
4	red	0.636 – 0.673	30	OLI
5	NIR	0.851 – 0.879	30	OLI
6	SWIR-1	1.566 – 1.651	30	OLI
7	SWIR-2	2.107 – 2.294	30	OLI
8	pan	0.503 – 0.676	15	OLI
9	cirrus	1.363 – 1.384	30	OLI
10	TIR-1	10.600 – 11.190	100	TIRS
11	TIR-2	11.500 – 12.510	100	TIRS



https://blogs.fu-berlin.de/reseda/landsat-8/

Random forest parameterization

```
var classifier = ee.Classifier.smileRandomForest({numberOfTrees:10})
```

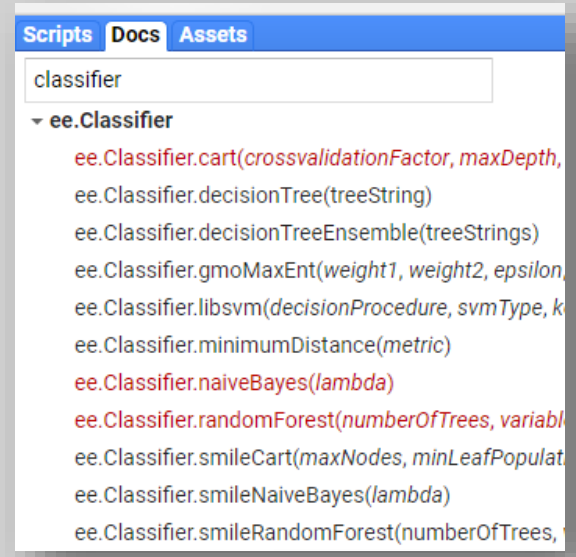
Other arguments:

```
ee.Classifier.smileRandomForest(numberOfTrees, variablesPerSplit,
minLeafPopulation, bagFraction, maxNodes, seed)
```

Other classifiers:

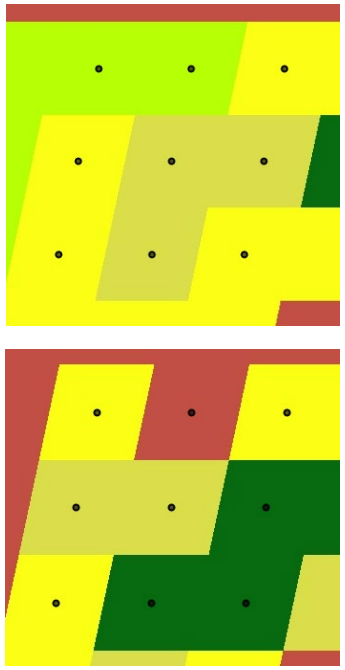
```
ee.Classifier.smileNaiveBayes
ee.Classifier.smileCart
ee.Classifier.libsvm
```

```
ee.Clusterer.wekaKMeans
```
← unsupervised clusterer



LULC Confusion Matrix

OA = 33.3%




```
2: [0,0,0,0,0]
8: [2,1,0,0,0]
9: [1,1,2,0,0]
10:[0,0,1,0,1]
12:[0,0,0,0,0]
```

Predicted

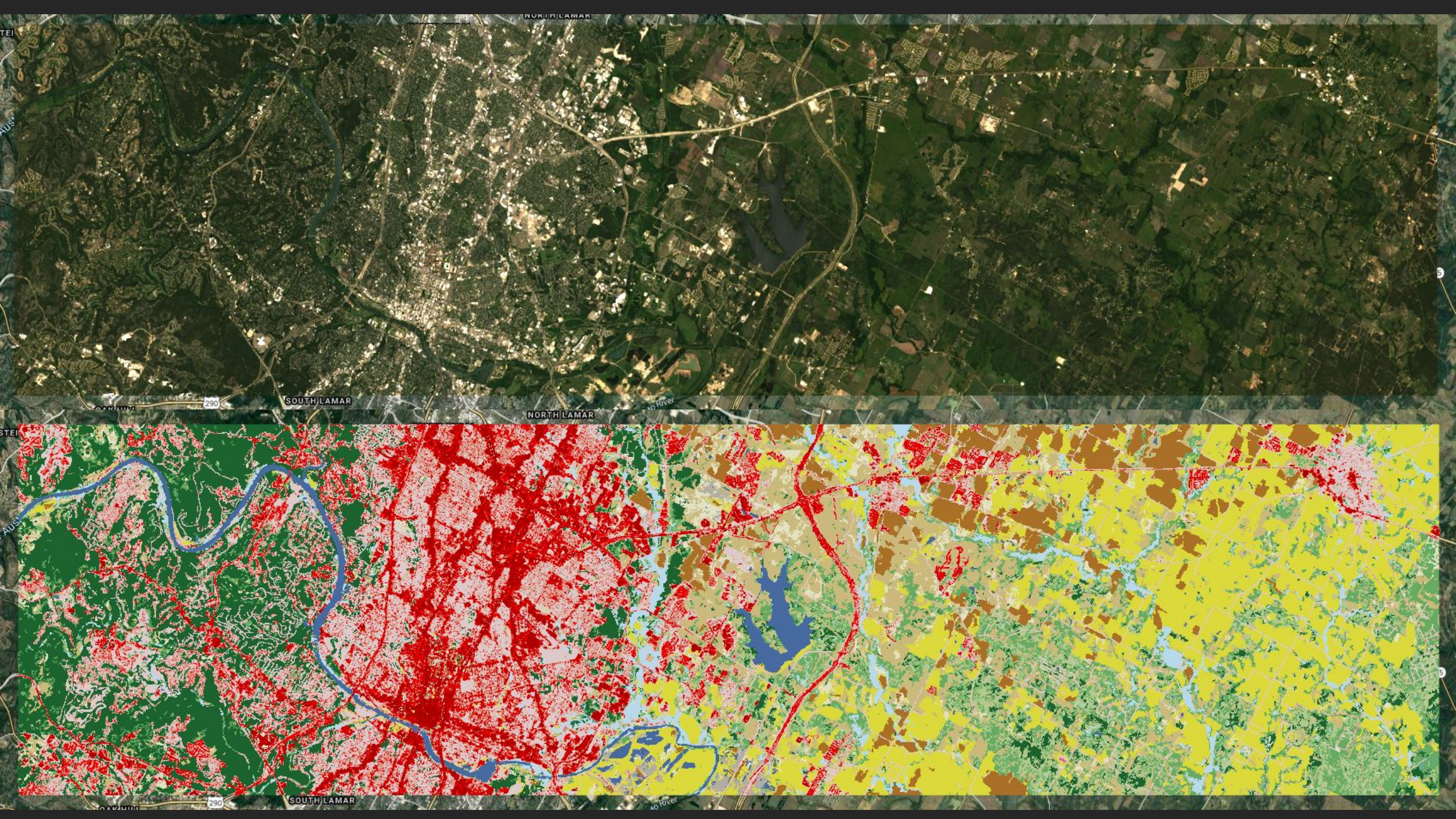
	Forest	Woody	Savan.	Grass	Agri.	TAct	Loss
Forest (green)	0	0	0	0	0	0	0
Woody (beige)	2	1	0	0	0	3	1
Savan. (yellow)	**1**	**1**	2	0	0	4	2
Grass (lime)	0	0	1	0	**1**	2	2
Agri. (red)	0	0	0	0	0	0	0
TPred	3	2	3	0	1		
Gain	3	1	1	0	1		
Net Diff	3	-1	-1	-2	1		

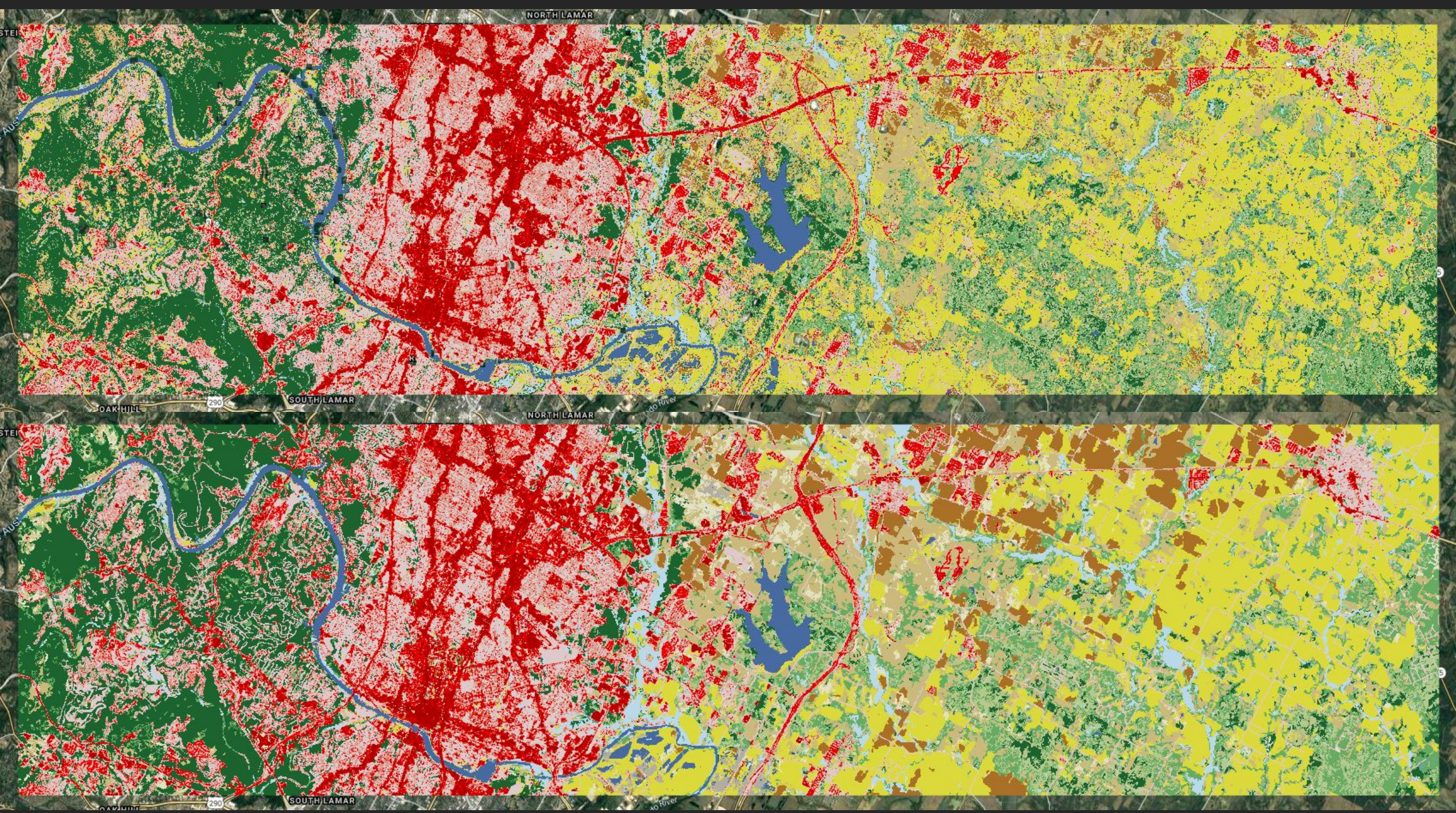
Actual

- **Savannas** (yellow)
- 4 pixels in actual
- 3 pixels in predicted
- 2 pixels were lost
- 1 pixel was gained
- Net difference of -1
- 1 savanna pixel was misclassified as woody
- 1 pixel savanna pixel was misclassified as forest

How many woody pixels were misclassified? [2]

How many grass pixels were misclassified as agriculture? [1]


NORTH LAMAR
SOUTH LAMAR
290
NORTH LAMAR
SOUTH LAMAR
290


NORTH LAMAR
SOUTH LAMAR
OAK HILL
290
NORTH LAMAR
SOUTH LAMAR
OAK HILL
290

Deliverables and evaluation

DUE 10/30/2020 by 6:00pm CST

Submit a lastnameFirstname_GEE_results.txt document that contains each of the required elements below.
Email solutions to ML4SCI2020@gmail.com

- Link to GEE script (use the 'Get Link' button at the top of your code editor)
- Bands selected for parameterization
- Classification method and parameters selected
- Sample size selected
- Rationale for each of the parameters set or post hoc explanation
- Highest overall training/validation accuracy
- If you used a method other than what was supplied here, please detail your methods, parameters, and results, otherwise put N/A here.

Evaluation will be based primarily on the overall accuracy achieved using the validation sample, but consideration will also be given to the explanation of parameterization.

- Overall accuracy calculated from a confusion matrix