Alteia

Agriculture Machine Learning

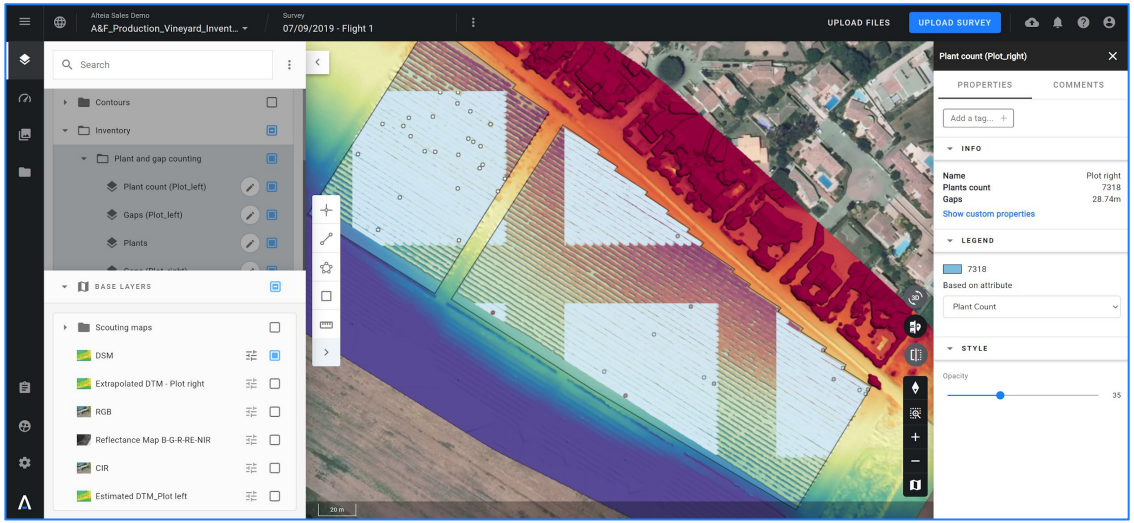
Who am I

- Head of Data science team @ Alteia 💾
- Passionate about Deep learning and math in general 🤖
- You can contact me here: Linkedin 👋

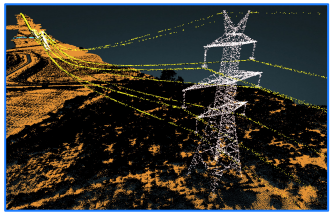
Alteia

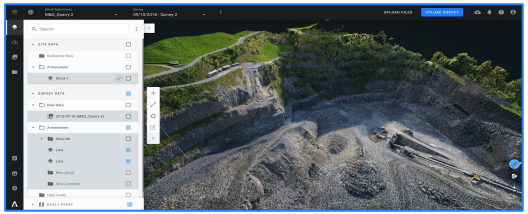
MANAGE PHYSICAL ASSETS - 2D▪3D▪4D

- Run Analytics from satellite
- Extract Insights from drone mapping
- Model Reality from air & ground scanners
- Inspect & Detect from mobile devices









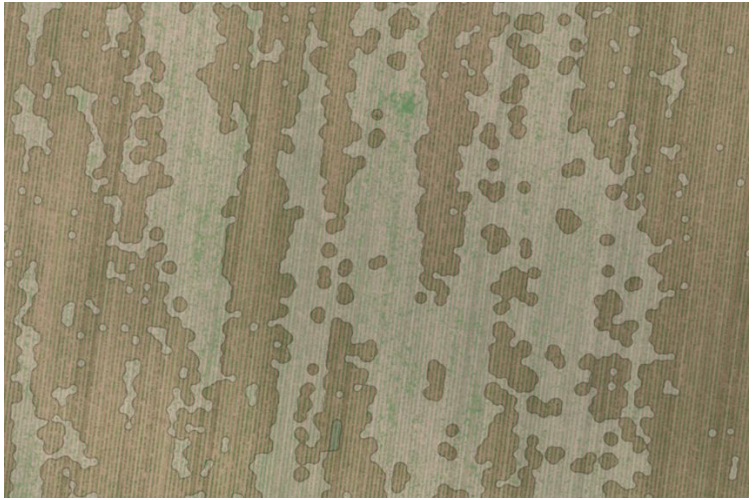
Agriculture and Data Science

Potato Healty

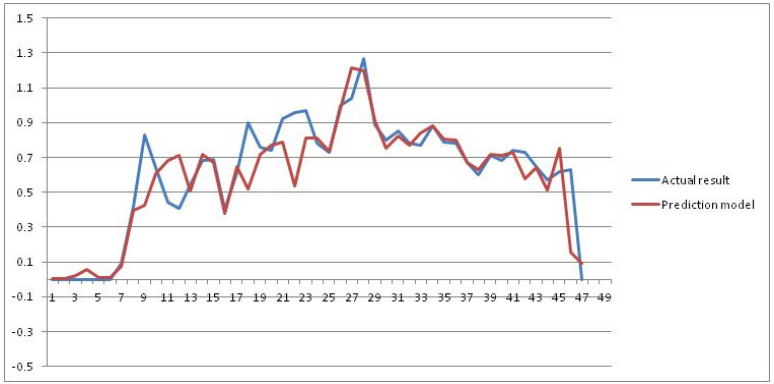
Potato Early Blight

Tomato Healty

Tomato Bacterial Spot



- Yield prediction
- Disease detection
- Weed detection
- Crop recognition
- Crop quality



Plan

- Generalities of Machine Learning
- Data acquisition methods and data sources
- Case study
- Conclusions and Q&A

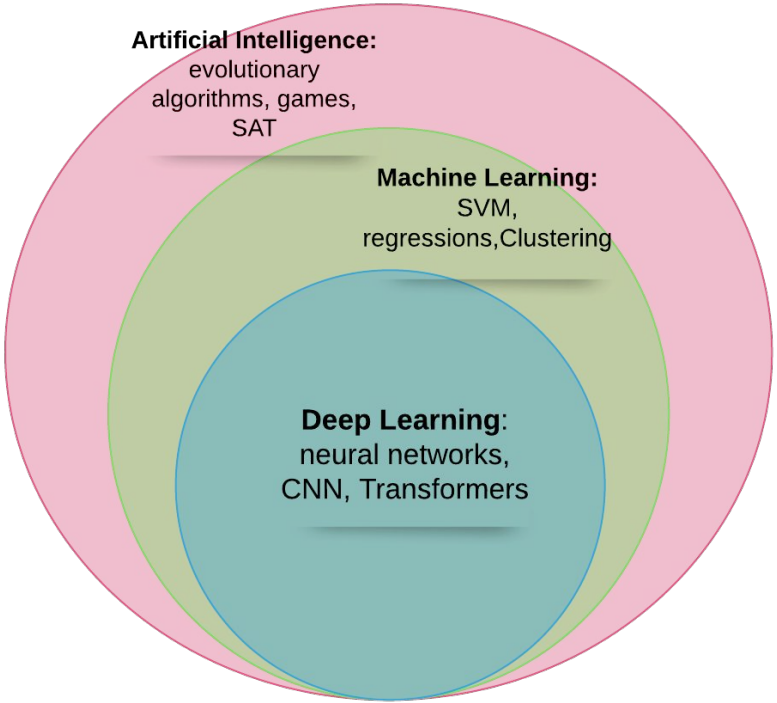
Machine Learning

Artificial Intelligence de-buzz the buzz word

AI is just a set of techniques to solve some technical problems.

It does not mean that is “intelligent” (what does it even mean ?)

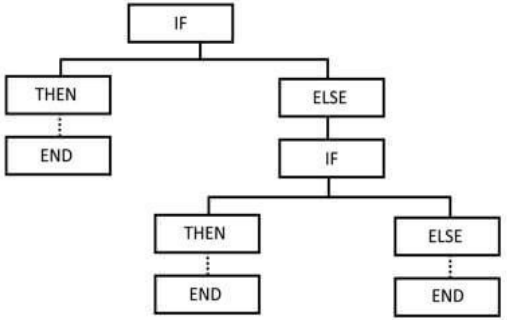
No robot army will rise in the near future. It is difficult enough to solve simple problems without any need of consciousness.



Machine Learning Programming model

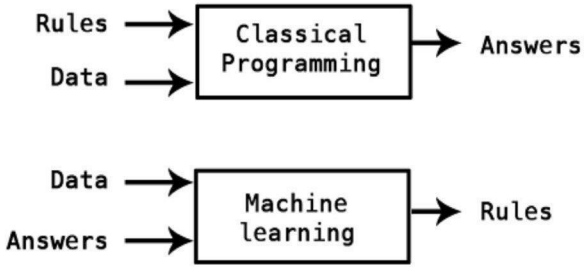
Rule-based: rules are explicitly given to the system

- Aims for decision-making
- Example: AIs in video games
- Con: Difficult to build such AI for complex problems

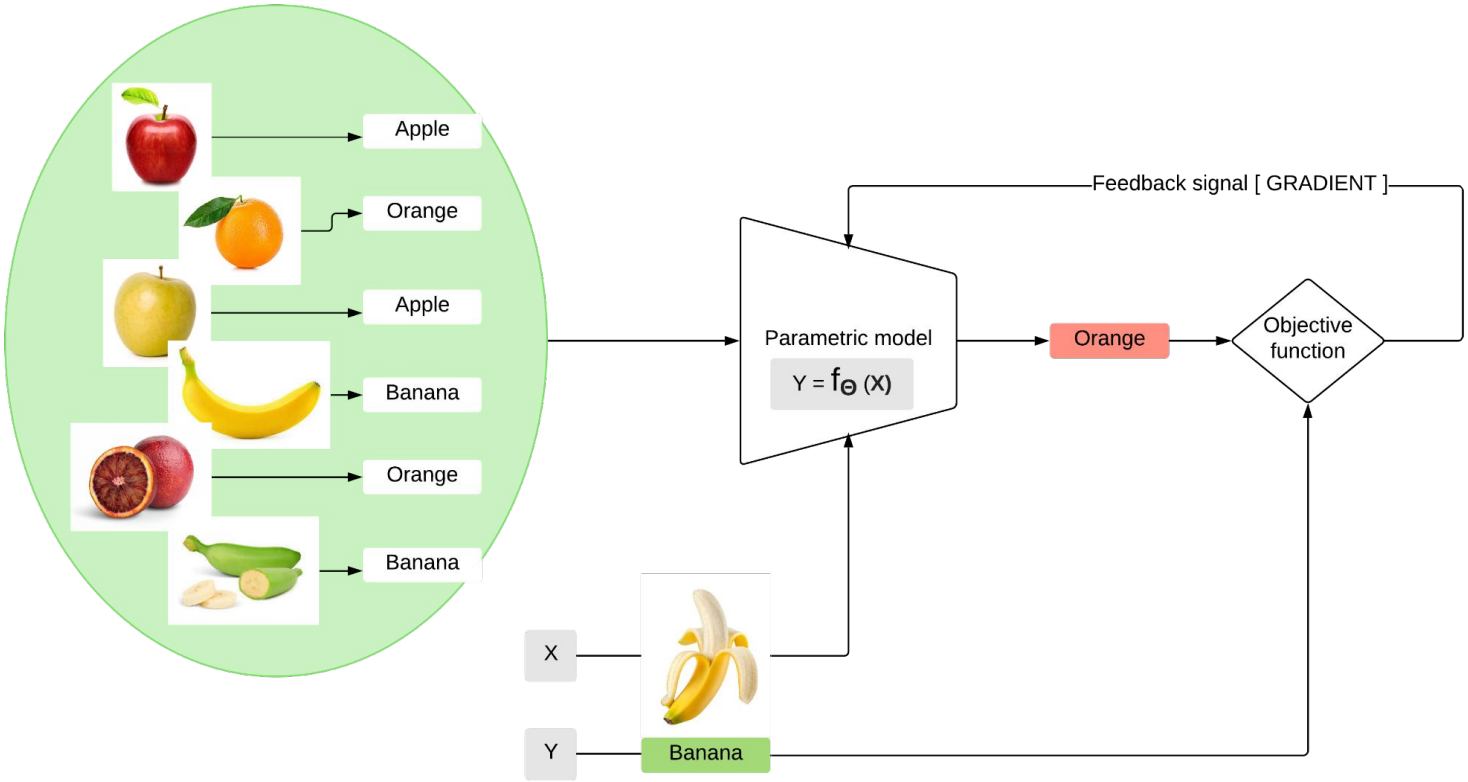


Data-based: problem inversion, the system learns the rules

- Finds patterns in the data
- Needs *A LOT* of data (Big Data)
- Can be hard to explain the outputs (black-box model)
- Machine learning is data-driven
- May have unpredictable behaviors



Key concepts in machine learning model training



Different paradigms

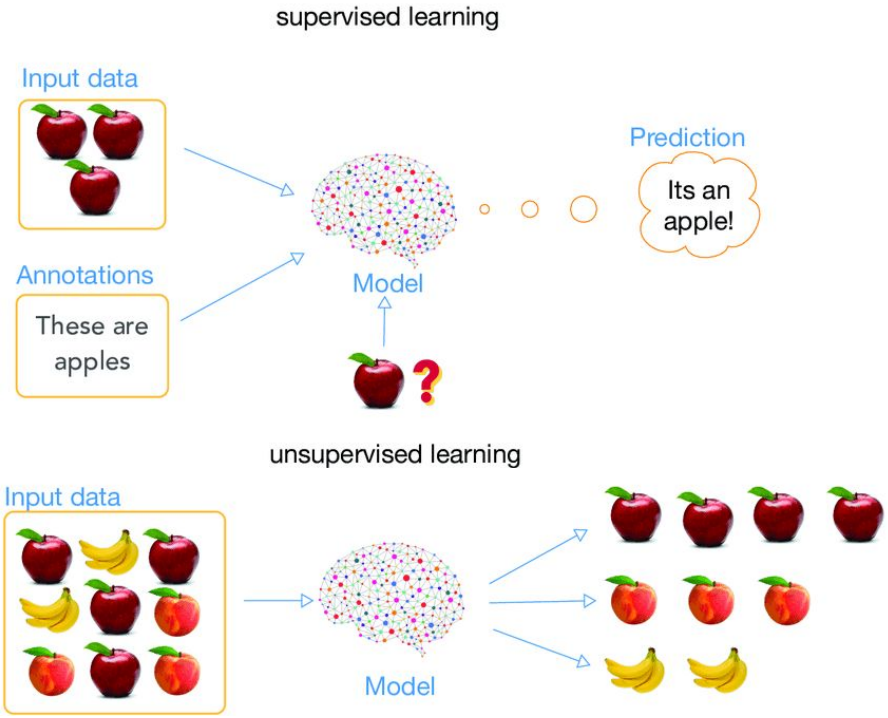
- *Supervised Learning:*
  - A data point is defined by its features and its label
  - Training on *labelled* data, predictions on new unlabeled data

- *Unsupervised Learning:*
  - A data point is defined by its features only
  - To separate *unlabeled* data into different categories

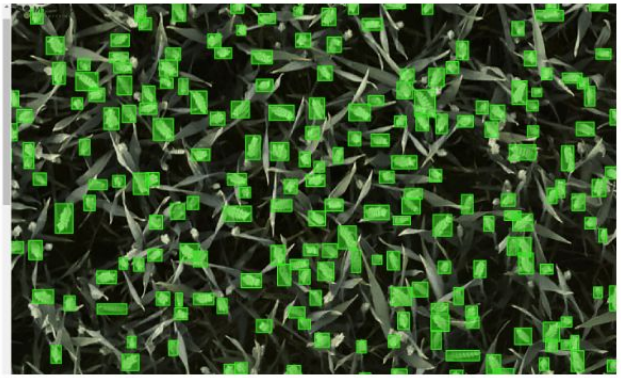
These are two different ways to solve the same problem.

The difference lays in the accuracy that we can get and the budget of data that we have.

Using unsupervised learning is less data efficient (usually).



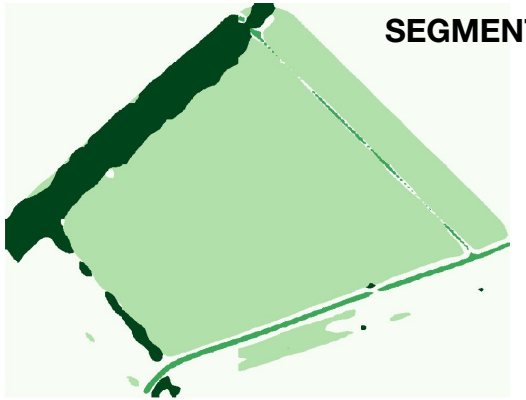
Agriculture tasks and problems



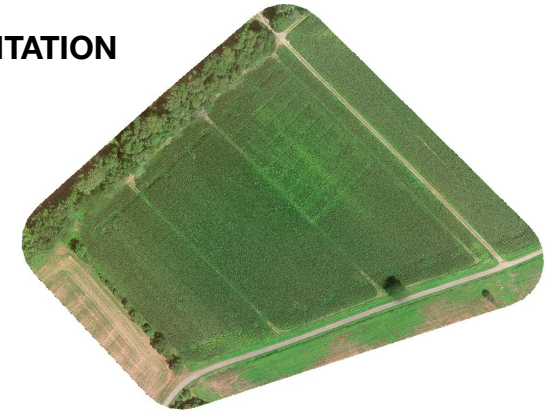
DETECTION

CLASSIFICATION

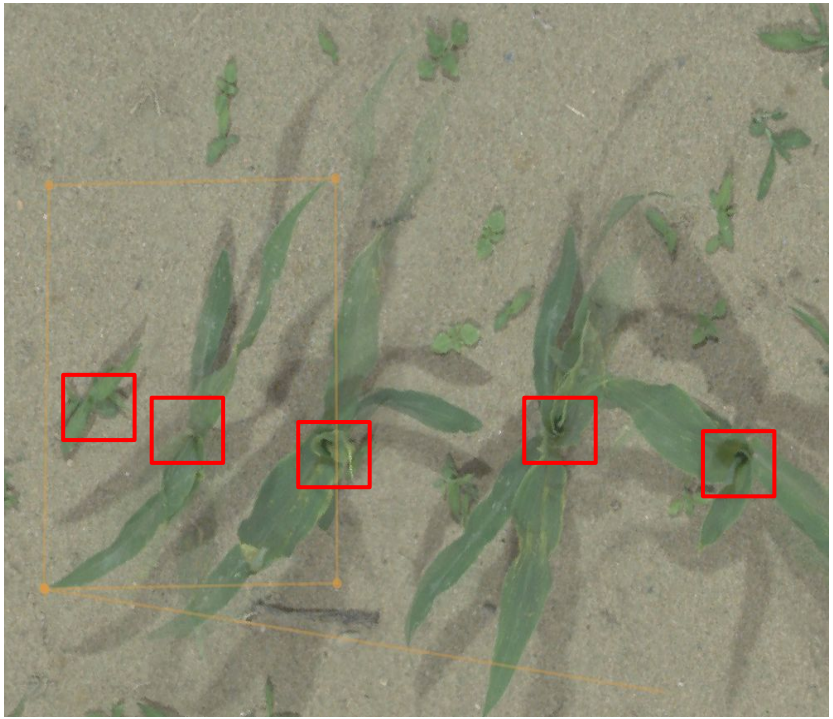
https://www.kaggle.com/c/cassava-leaf-disease-classification/data



SEGMENTATION

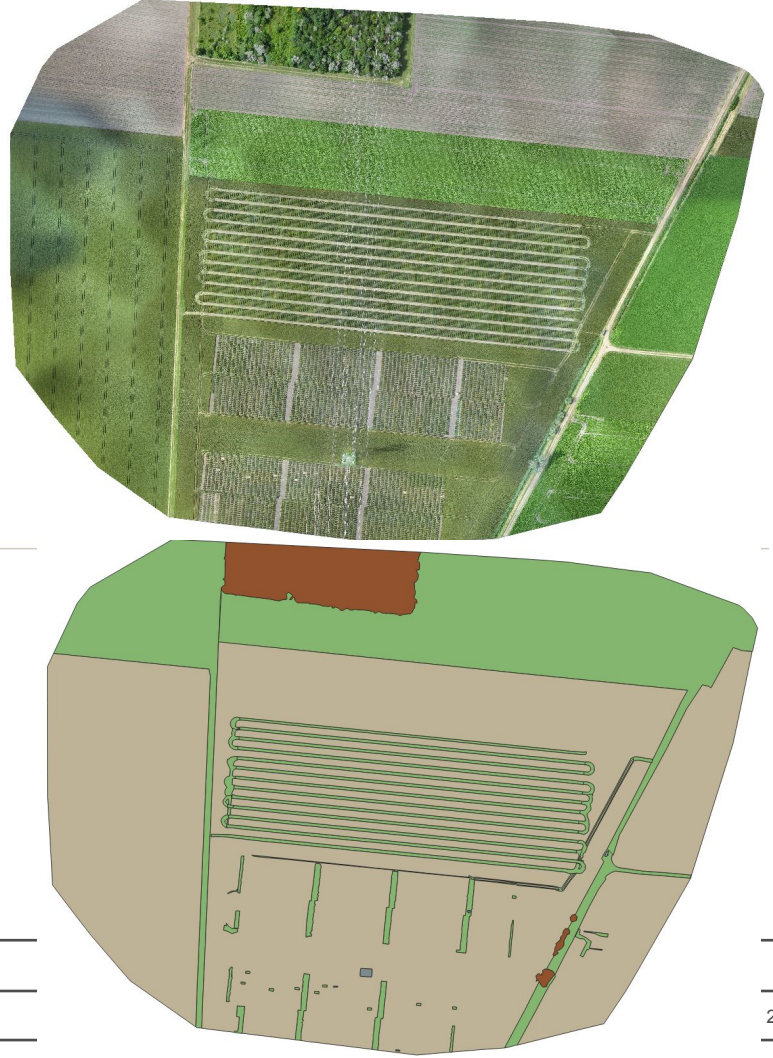


Importance of Labelling



Constrains of Labelling

- **Time:** do you need to label a dataset fast
- **Quality:** can you afford noise and errors in the dataset
- **Cost:** how much budget do you have
- **Location:** can you subcontract the label process
- **Type:** bounding box are easier than polygons



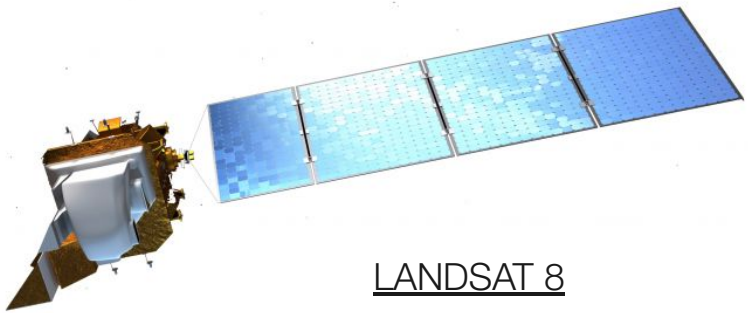
Data Acquisition

Remote sensing, AKA drones, robots and satellites

DELAIR

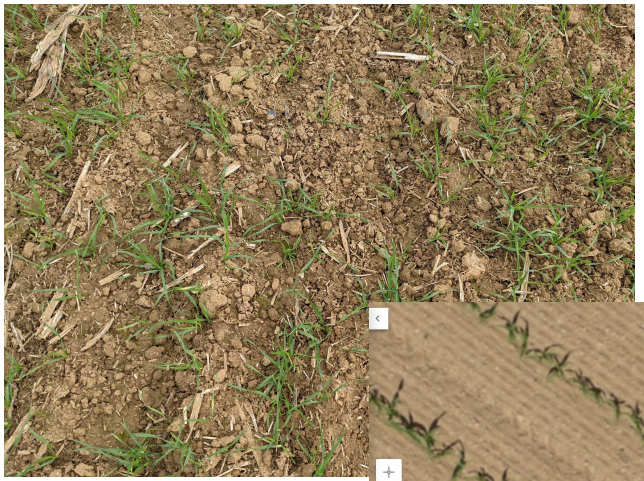
DJI

NAIO



LANDSAT 8

Spatial Resolution Importance

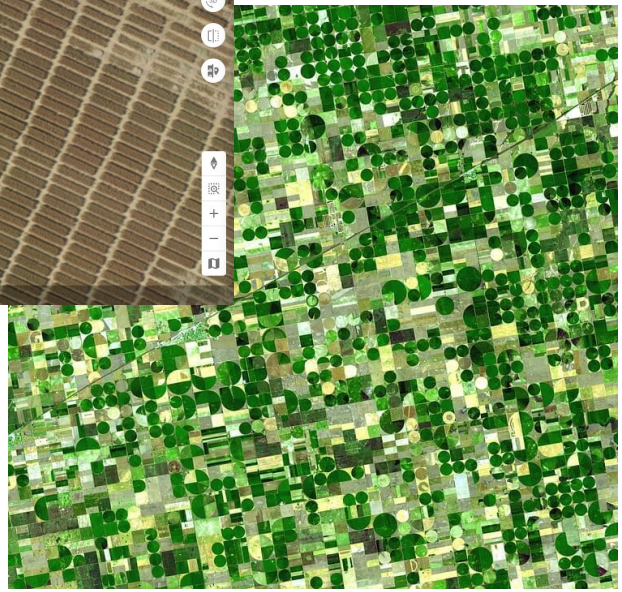


~ mm

~ 1 cm

~ 5 cm

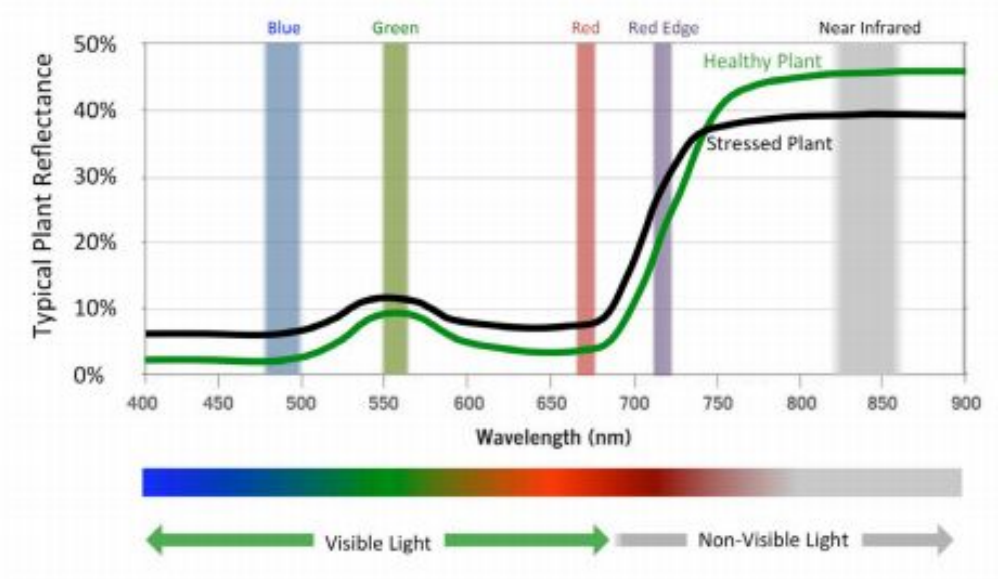
~ m



Cameras and sensors

Spectral bands

High-grade optical filters deliver precise information specially targeted to agricultural applications.

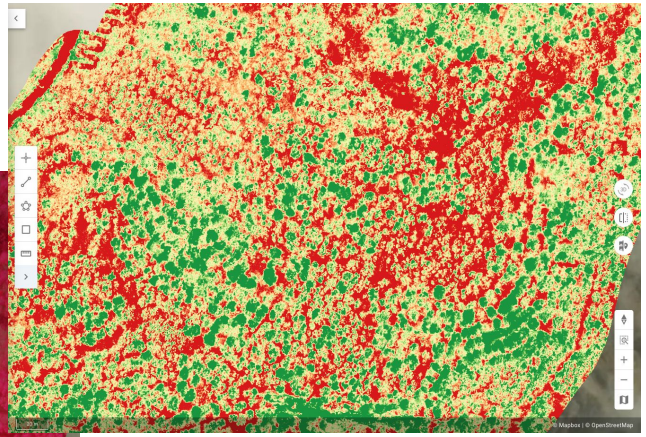


Micasense

Different kinds of data for different problems

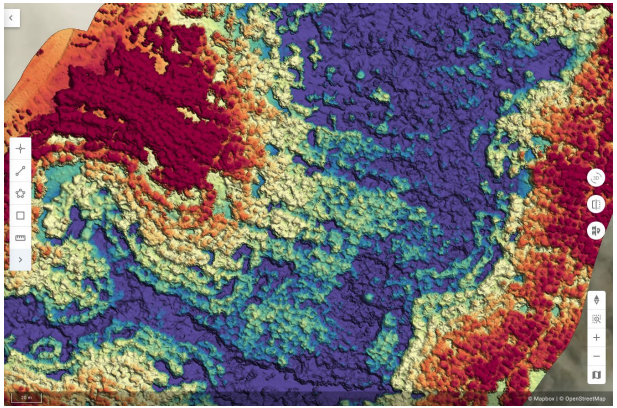
RGB

CIR

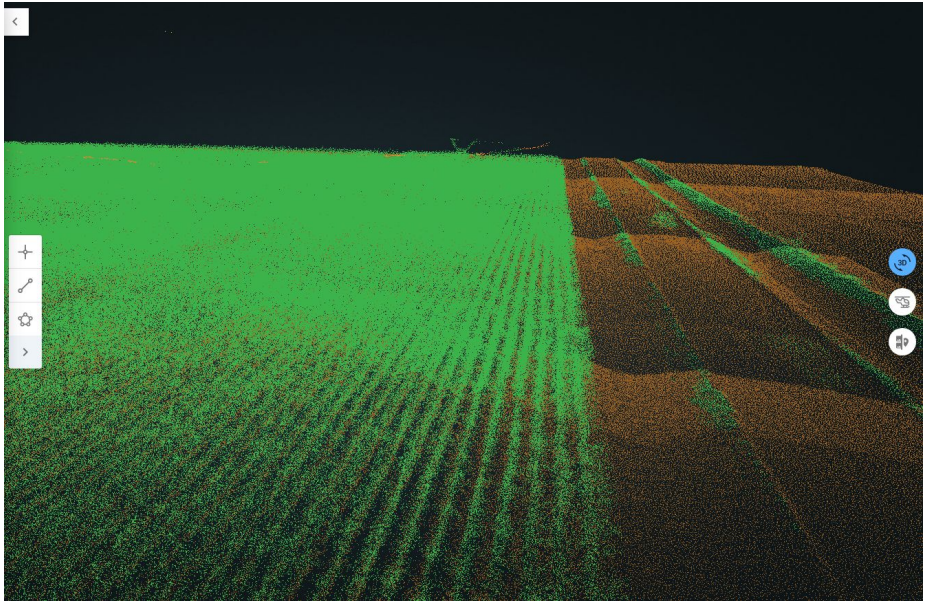


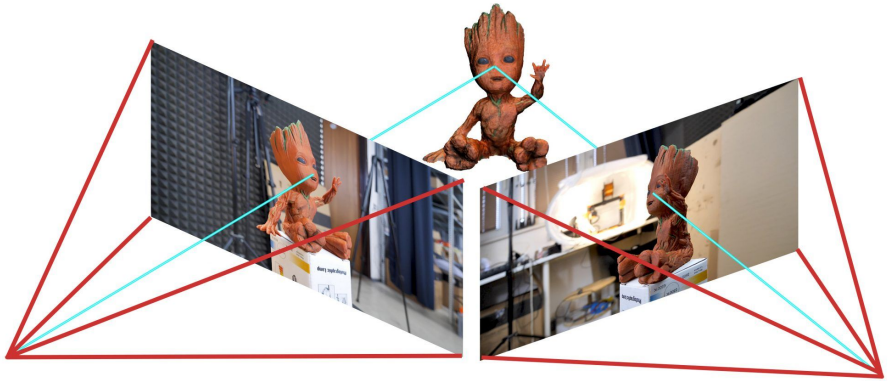
NDVI

3D Height Data



DSM





LIDAR

How much data do I need ?

BIG DATA is just a buzz-word:

You should ask yourself how much data do you need for a certain application and how much data you produce n a certain time frame; then select the solution of storage and retrieval accordingly

CLOUD

Pro:

- You do not manage bare metal
- You can pay for what you use
- Adapt to peaks and lows
- Always last technology

Cons:

- Can be difficult to setup
- Do not work offline
- Data retrieval speed can be slow

LOCAL

Pro:

- Works offline
- Data retrieval is faster
- Security (privacy politics)

Cons:

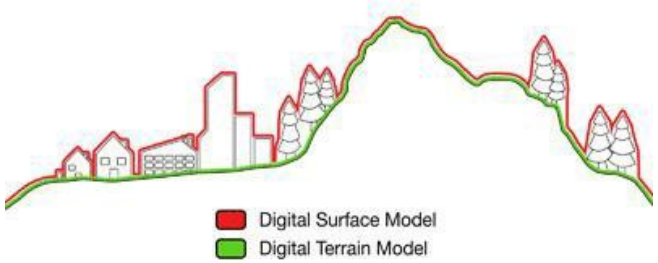
- You need to manage machines (install, update etc…)

CASE STUDY 1

Plant Height Estimation

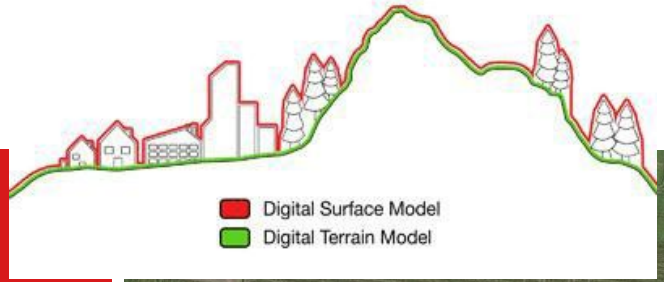
Objectives

- Estimate the height of the vegetation

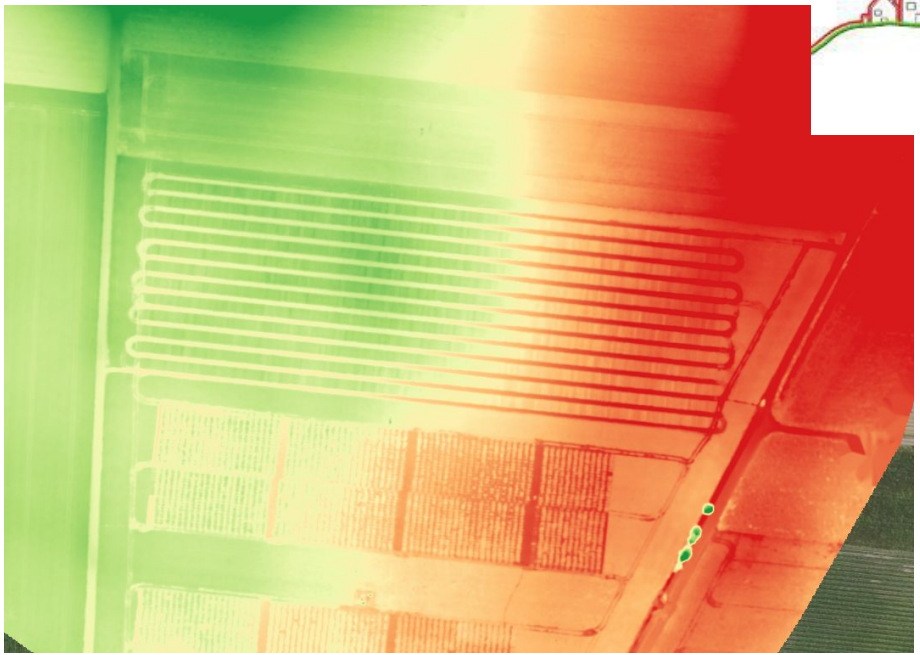


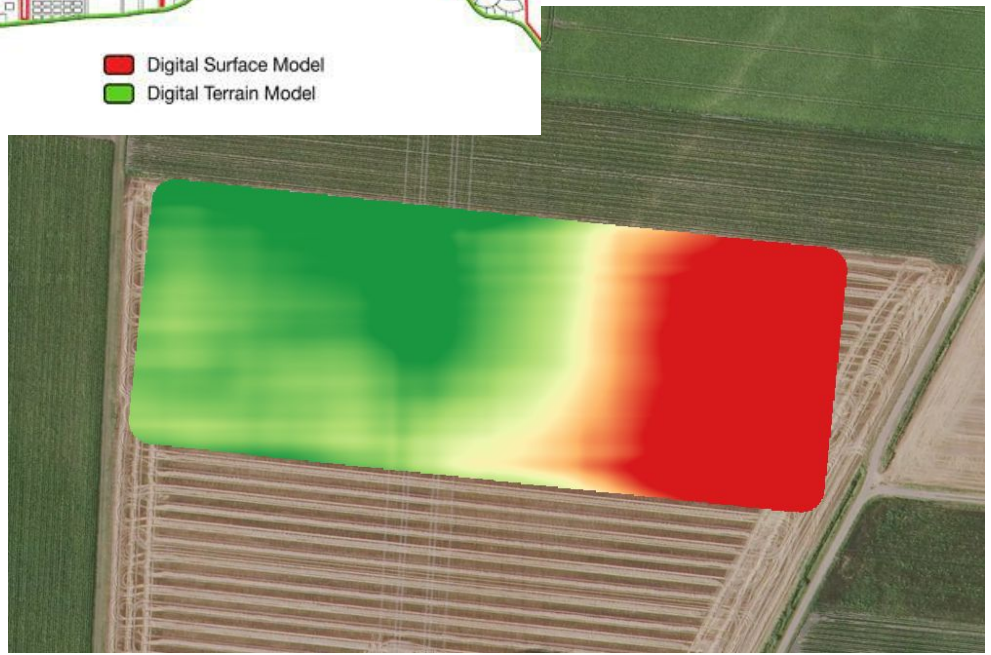
Dsm - Dtm

DSM

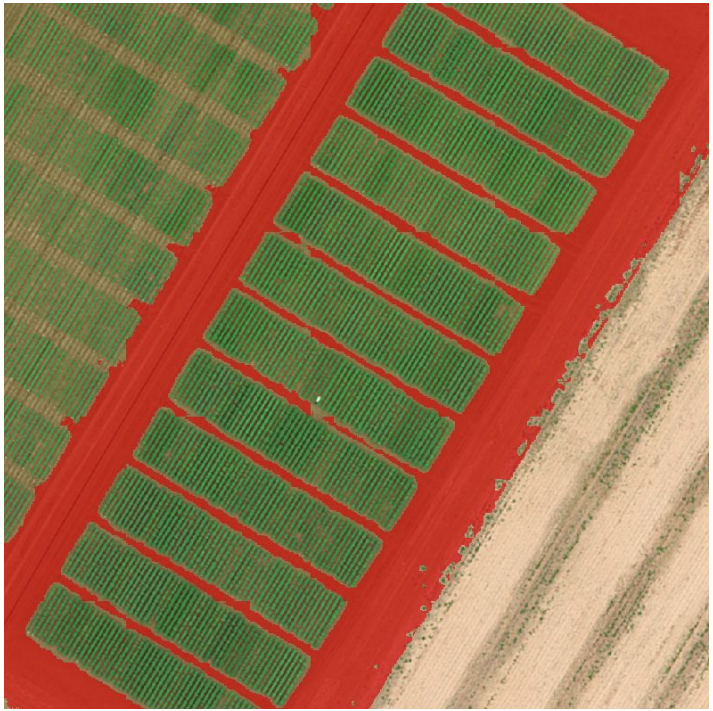


DTM



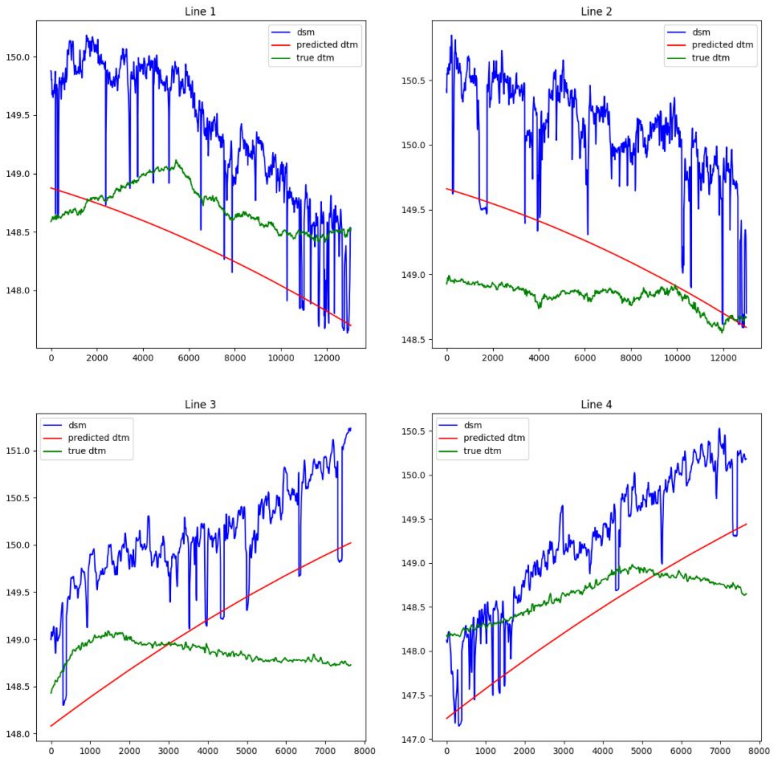


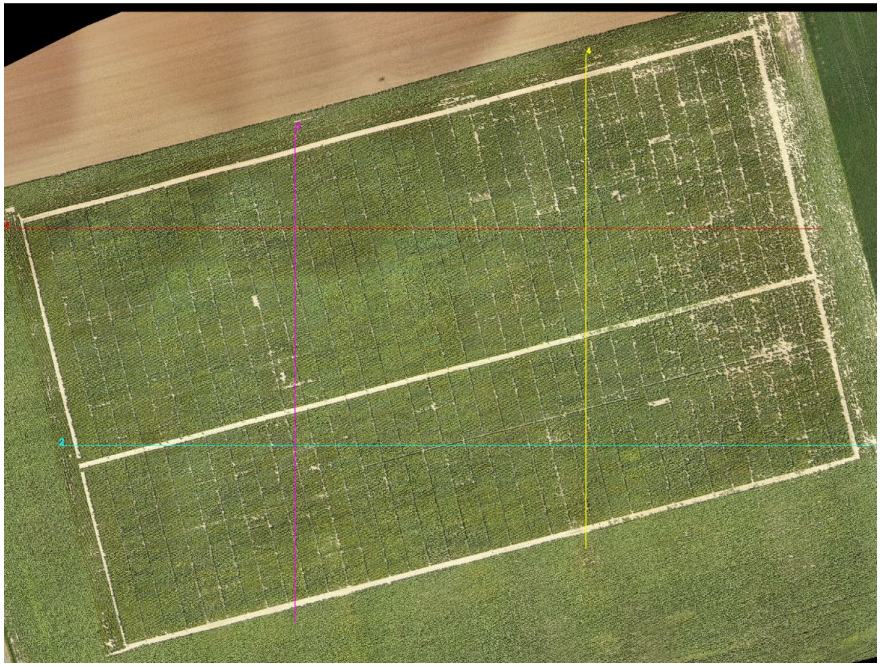
Ground detection



- LinkNet
  - Semantic segmentation on ground
- IoU is ~70% but it is not a problem

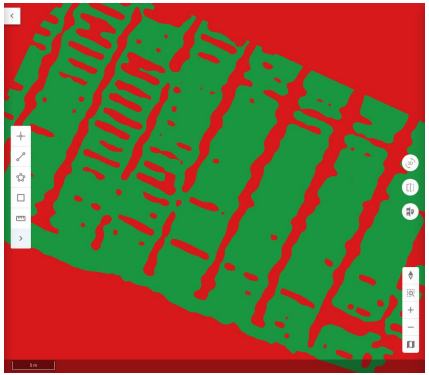
Results


Line 1
dsm
predicted dtm
true dtm
Line 2
dsm
predicted dtm
true dtm
Line 3
dsm
predicted dtm
true dtm
Line 4
dsm
predicted dtm
true dtm

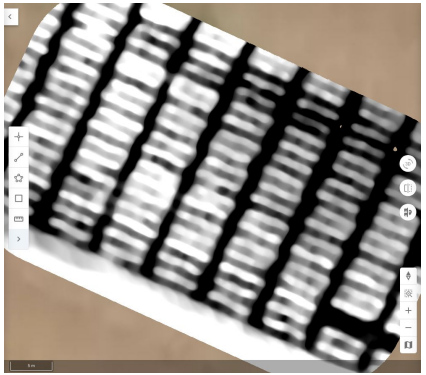


The big picture

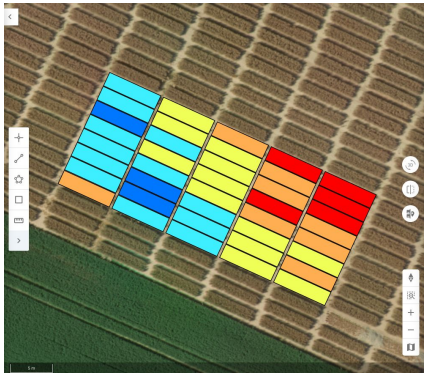
Capture: Raw Data



AI: Semantic Data



Digitalization: height model



Analytics: KPI

Our solution involves a multistep approach that transform progressively the raw data and each time provides new insights for the customer.

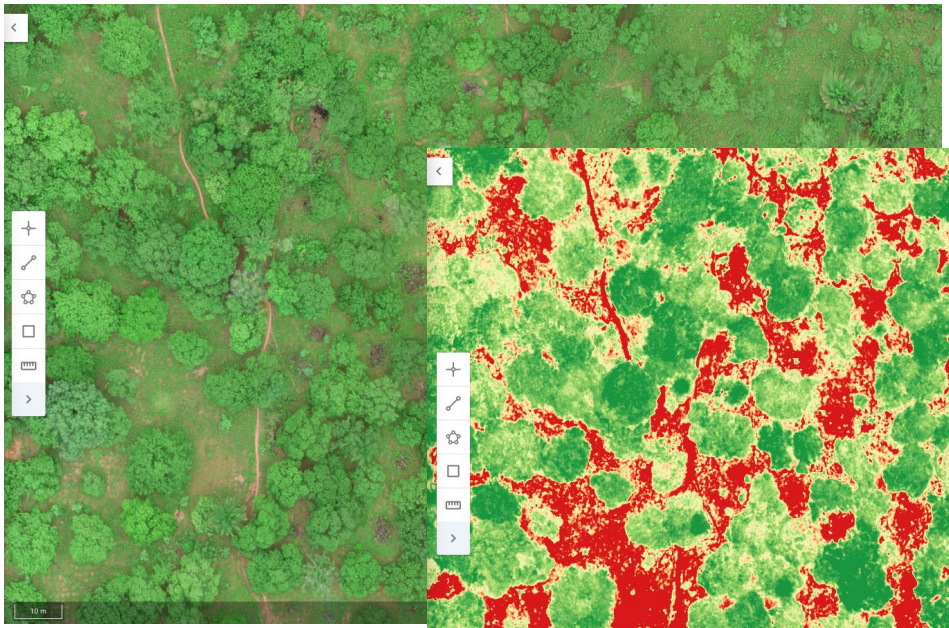
CASE STUDY 2

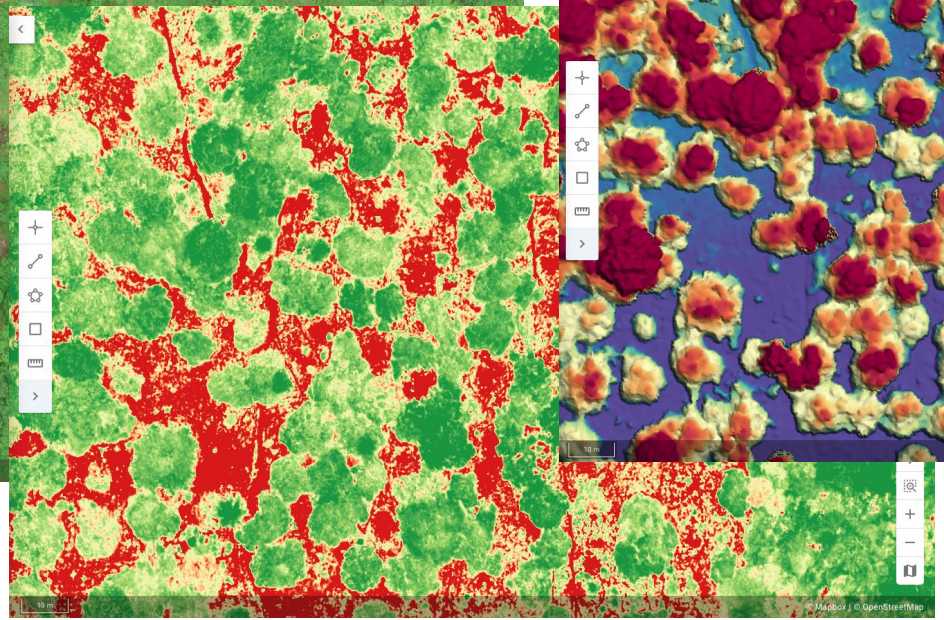
Tree Counting

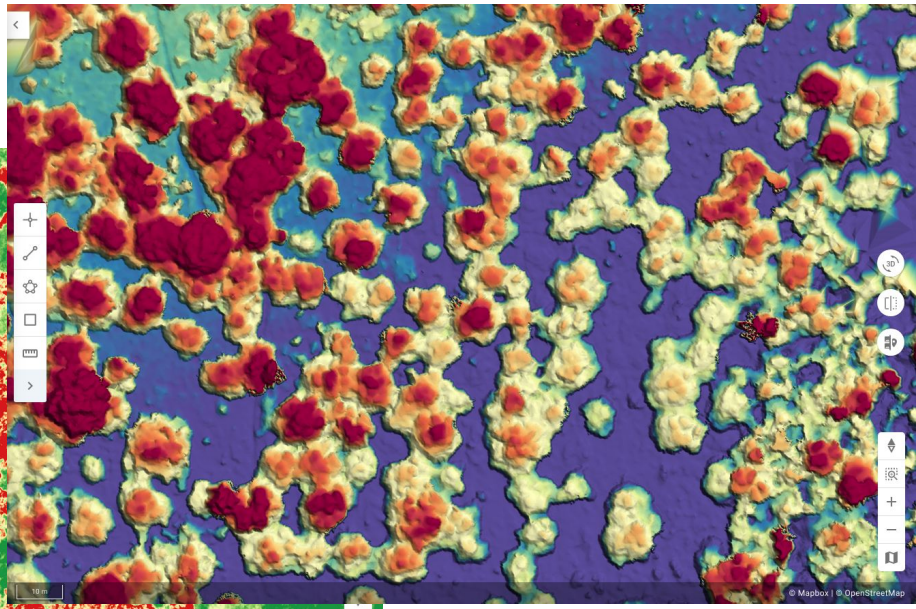
Forestry inventory

- Count how many trees there are in a certain area
- Define position, surface, height

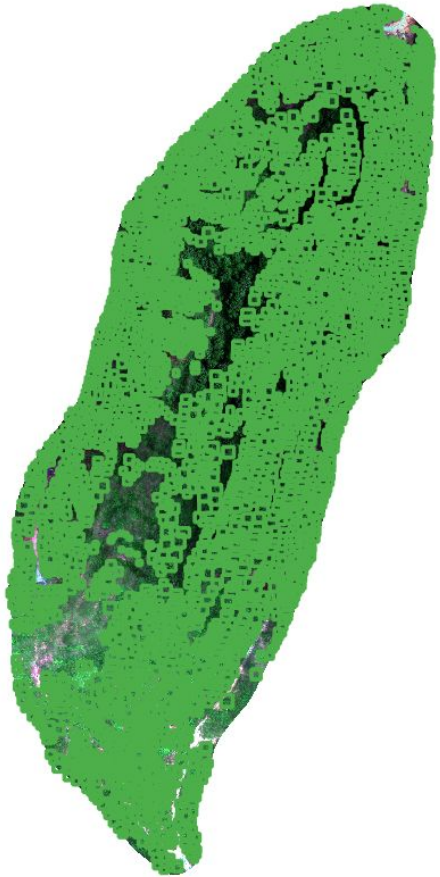
Data available

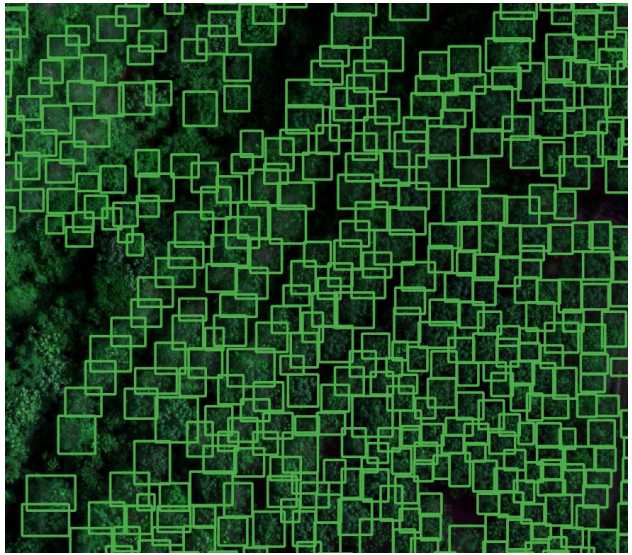


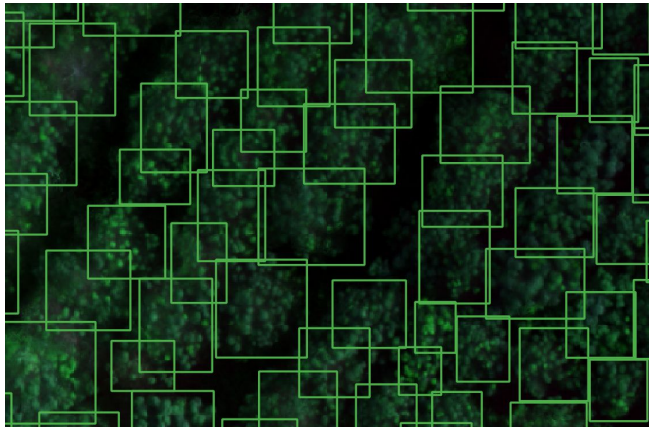




Bounding box labels

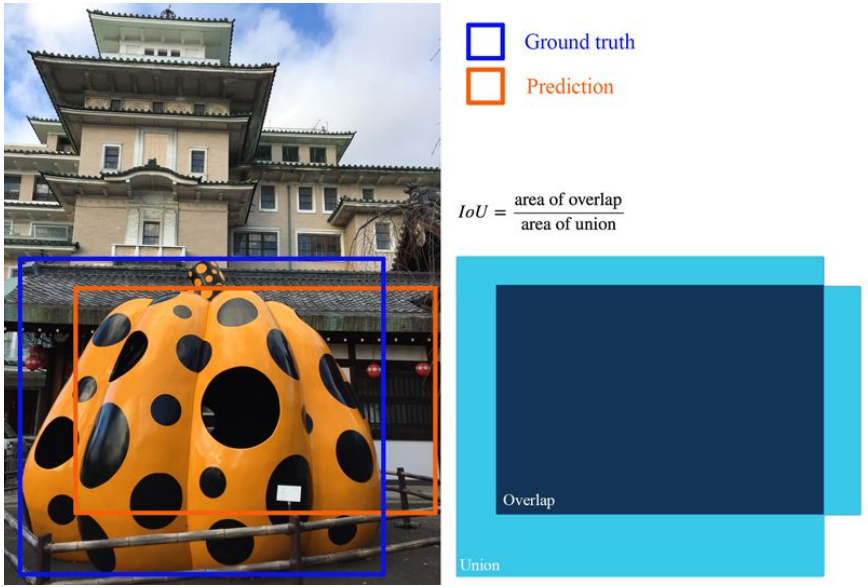






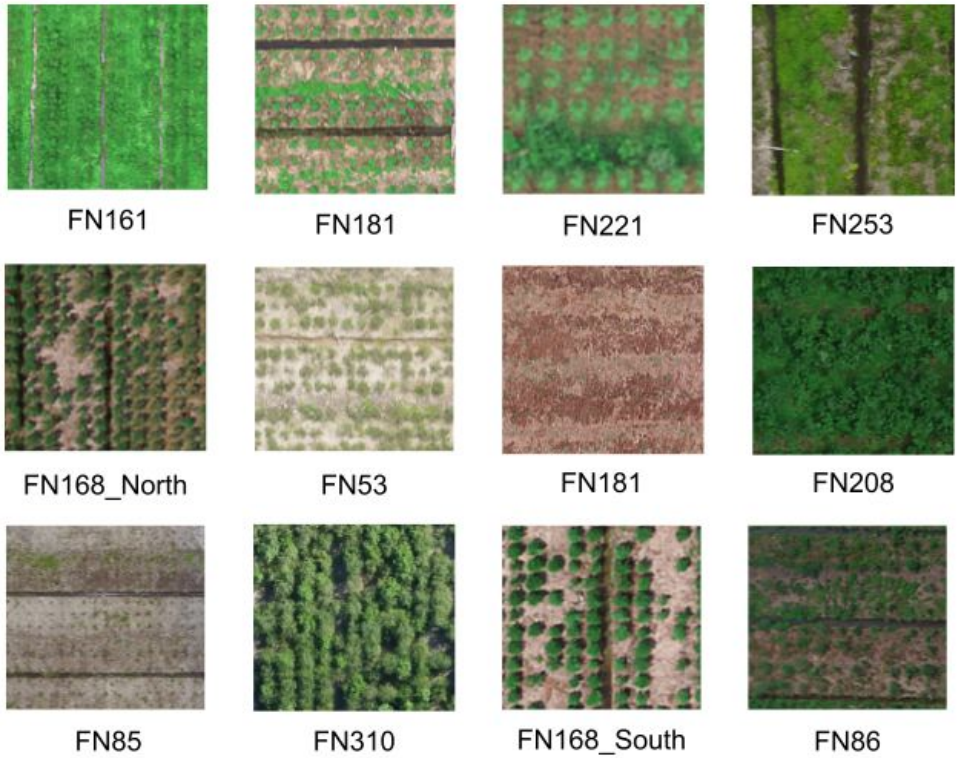
Supervised model

- Yolo
  - single stage detector
  - Fast and small
- Faster-RCNN
  - Two stage detector
  - Big and slow
- Uses multiple bands as input
- Metrics used:
  - mAP @ IoU = 50%
  - number of trees

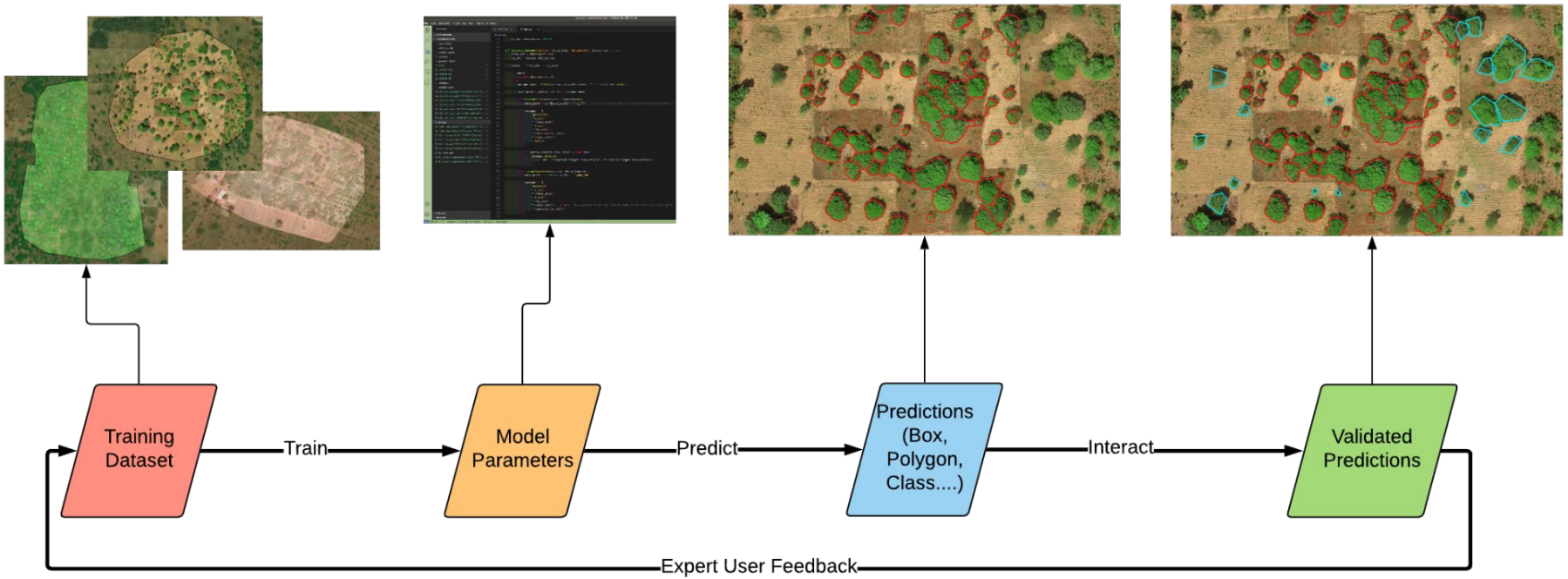


Some qualitative results

Domain adaptation



Interactive learning



Conclusions

Key take away

- One big mistake that everyone make is to think that humans are always 100% accurate. Measure, don’t guess !
- Fit the best **sensor - vector- algorithm - data** set with the problem you are trying to solve. The 4 are interconnected and not always easily interchangeable
- Data is more important than models (talk **From Model-centric to Data-centric AI**)
- The problem is more important than frameworks

Alteia Team

Gaston Lenczner
PhD student @Onera

Rémi Loze
Computer vision engineer

Lénaïc Grignard
Head of Agriculture and Forestry

Romain Laporte
Computer vision engineer

Hacene Karrad
Computer vision engineer

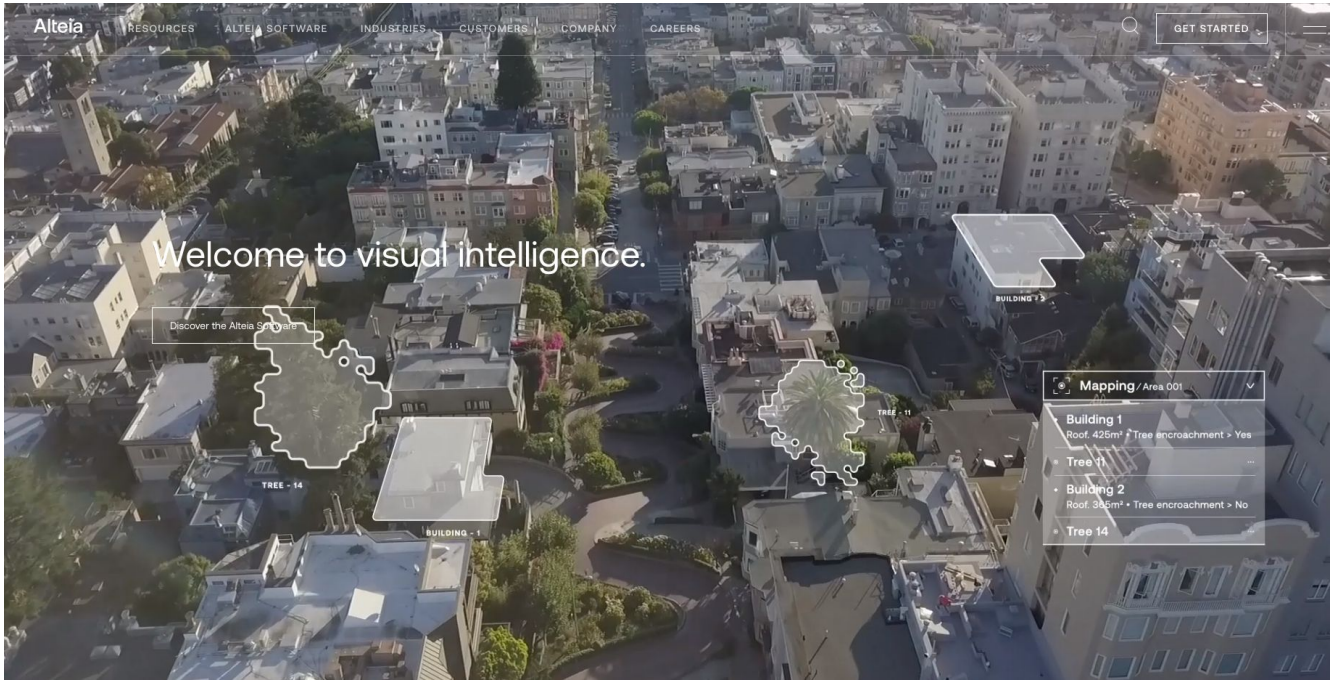
Quentin Vigneron
Agriculture Project manager

Aurélien Vila
Computer vision engineer

Nicola Luminari
Head of Data Science

Alexis Janson
Applied Product Engineer for Agriculture & Forestry

Data Scientist Wanted



[Join Our TEAM !]

Bibliography

- https://arxiv.org/abs/1506.02640
- https://arxiv.org/abs/1506.01497
- https://arxiv.org/abs/1707.03718
- https://www.researchgate.net/publication/227304926_Kriging_interpolation_methods_in_geostatistics_and_DACE_model
- https://github.com/davemlz/awesome-spectral-indices
- https://www.kaggle.com/c/global-wheat-detection
- https://alicevision.org/#photogrammetry


NOW I WILL TAKE
YOUR QUESTIONS.


THANK YOU!