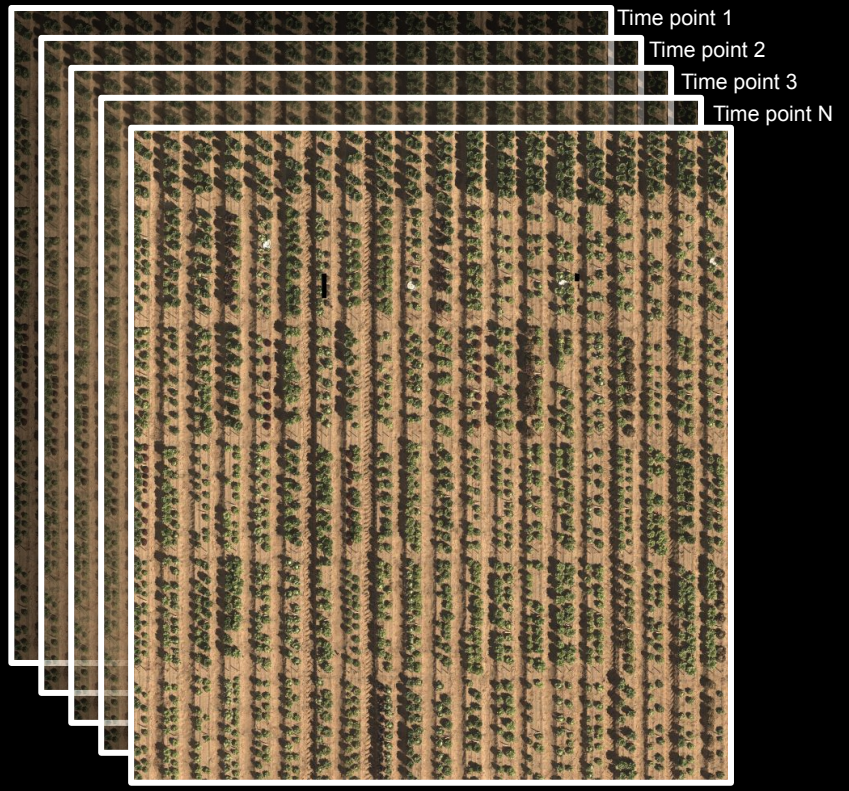
As easy as 1-2-3: Machine learning in real-world projects

ACIC 2021

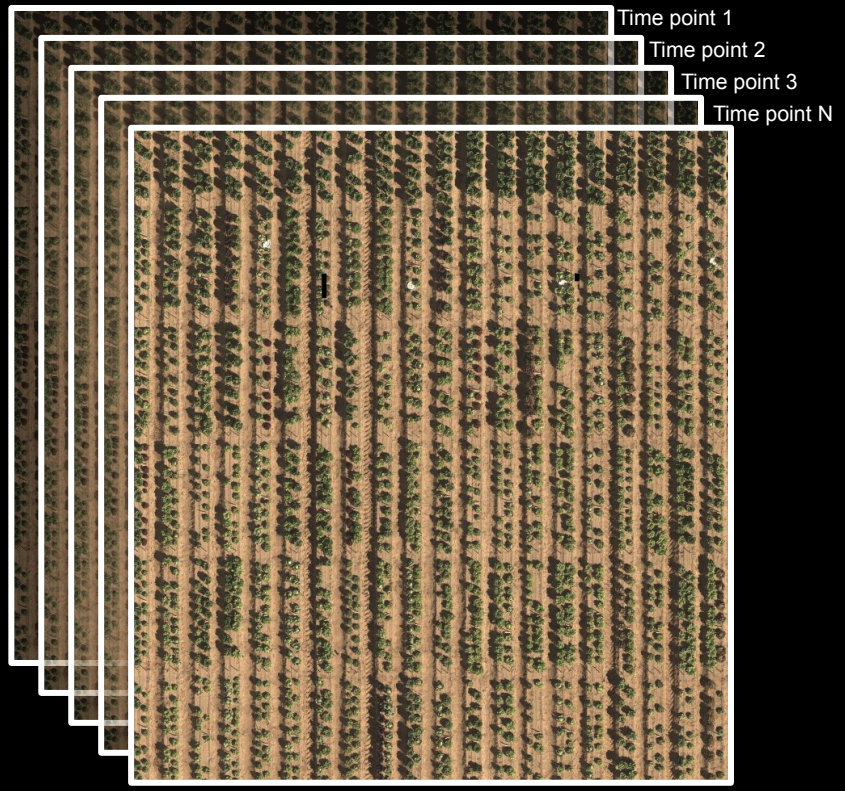
Problem Framing: Tracking thousands of plants

- Plant Science perspective
  - Interested in assessing phenotypic variation
    - Can we use our data to predict yield ?
    - Can we use our data to improve crop performance?
  - What other data and/or ML approaches will allow us to understand plant growth?

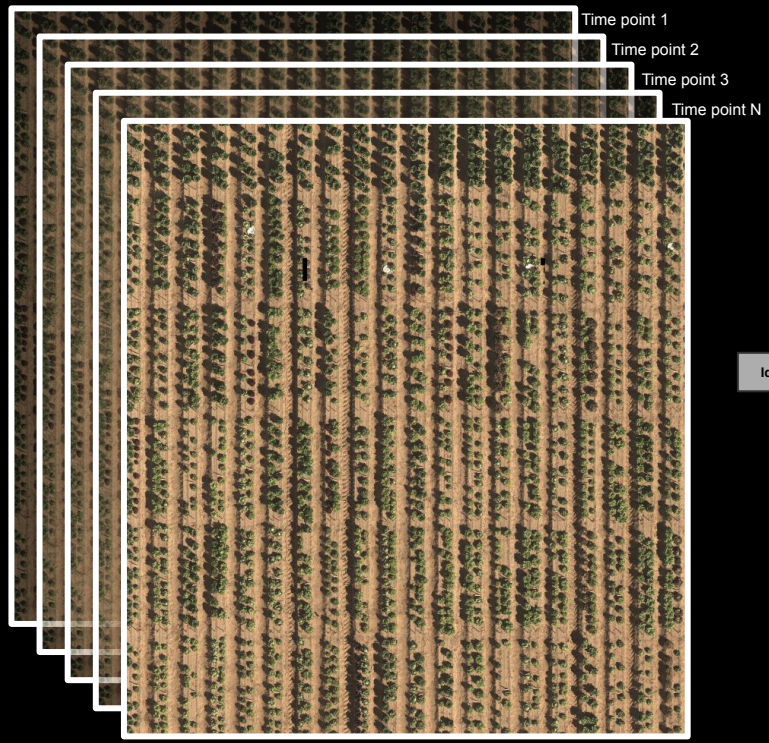


Problem Framing: Tracking thousands of plants

- Data Science perspective
  - Multiple data time points throughout growing season
    - Lots of training data = good!
  - Image processing techniques have been unsuccessful
    - Shift to ML methods



Problem Framing: Tracking thousands of plants



Ideas

Team

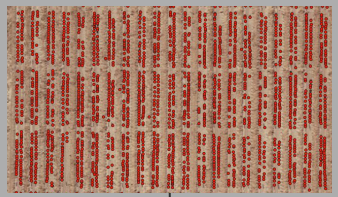
- Engineering
- Data Science
- Plant Science
- Computer Science

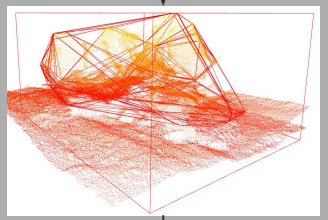
Sweat

Years

Tears

Team goals!



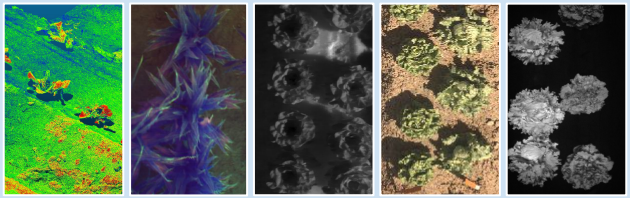


Discovery!

Data Collection: Converting raw data into gold

Collection method

Data management

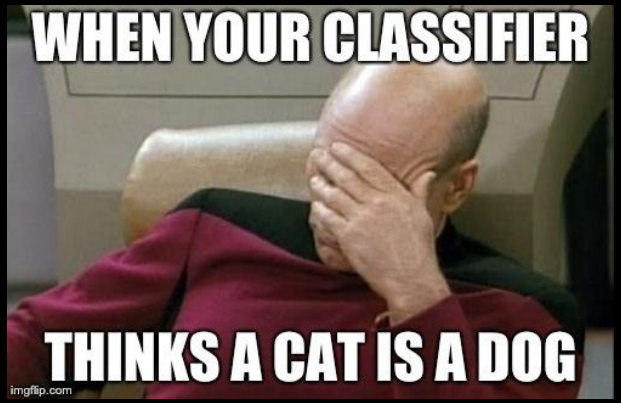


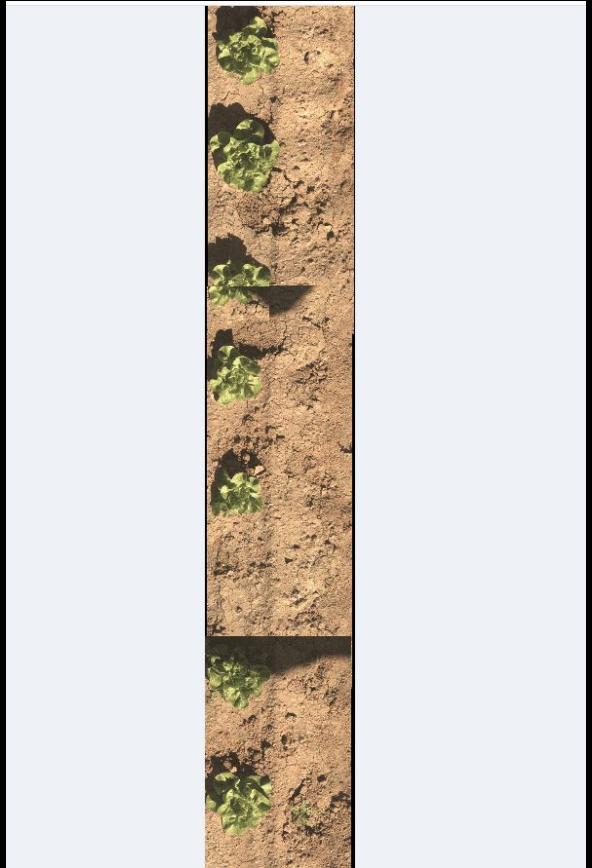
Data volume:

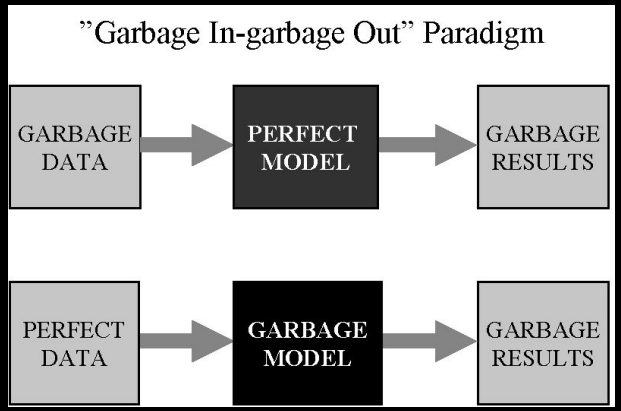
- Max: 10 TB/day
- Typical: 1.5 TB/day

Data curation

Data Preparation: Garbage in, garbage out

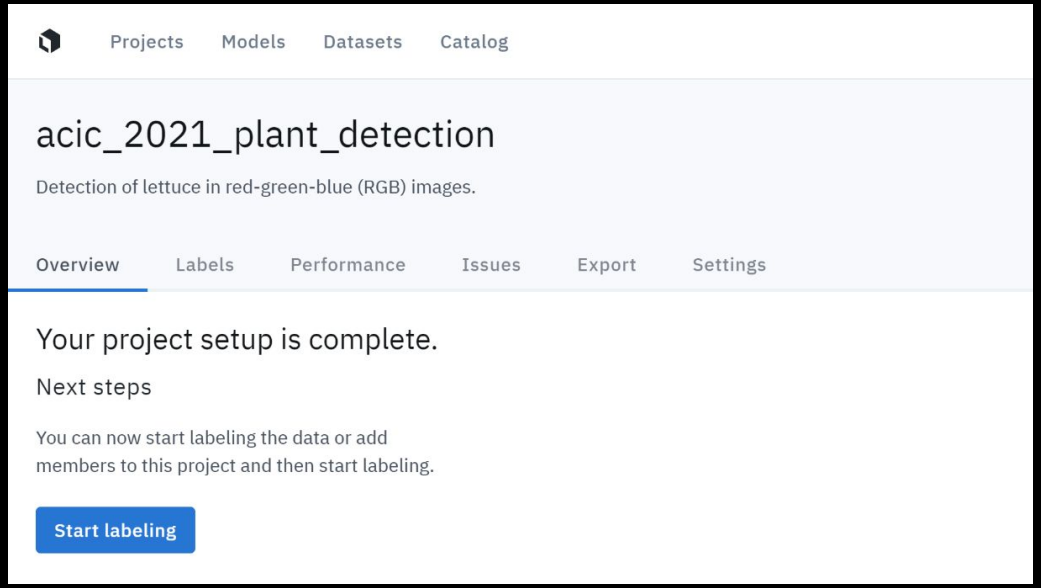


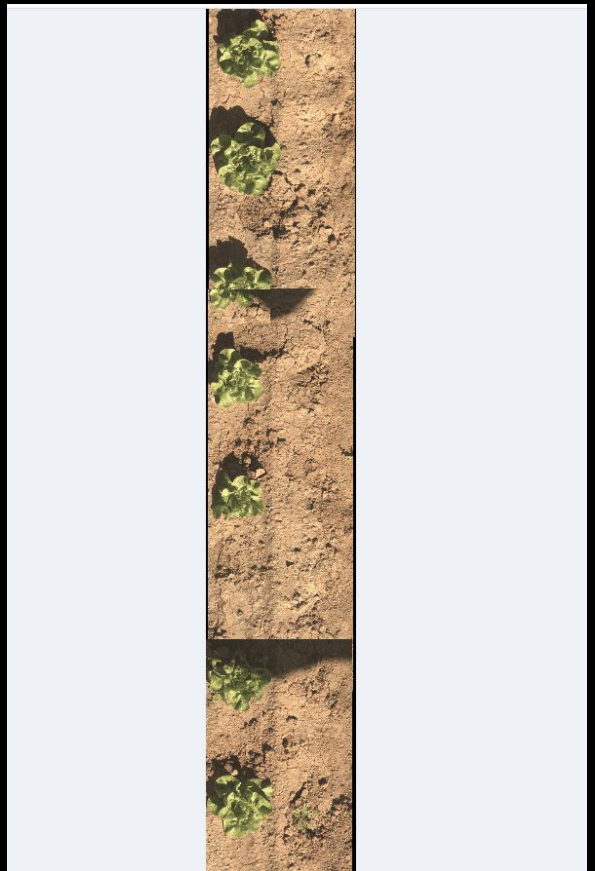




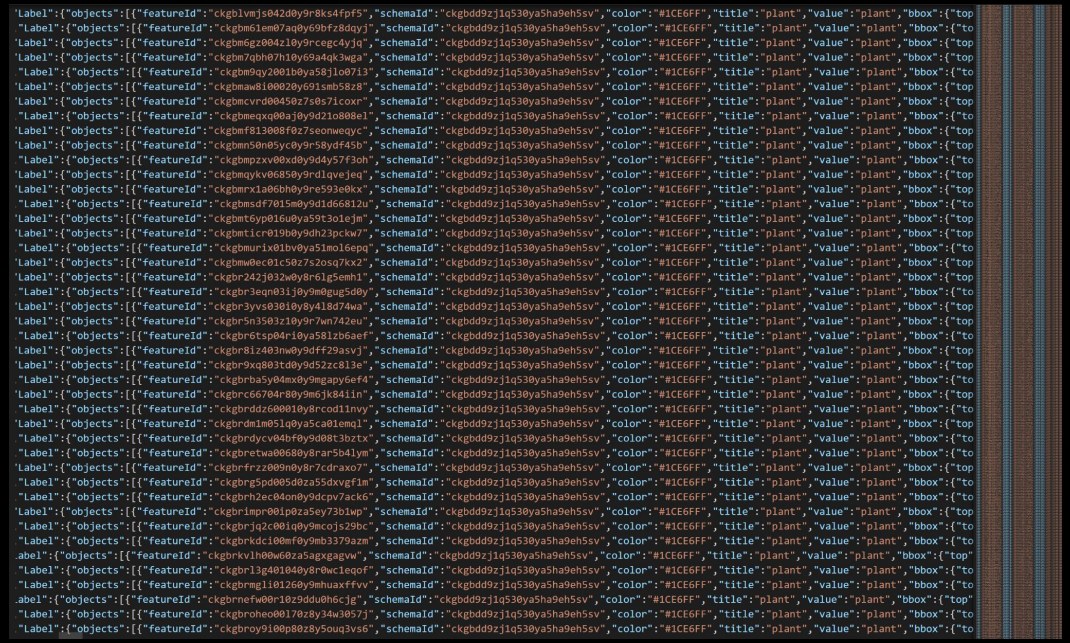
Data Preparation: Garbage in, garbage out

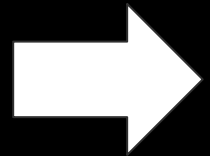
Let's label some images: [Labelbox]

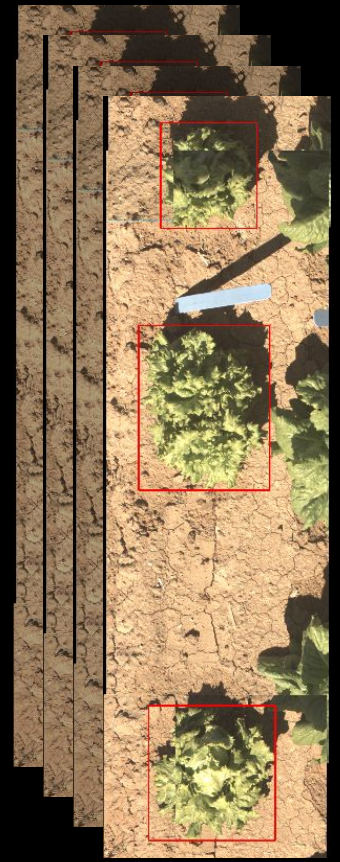




Data Preparation: Garbage in, garbage out



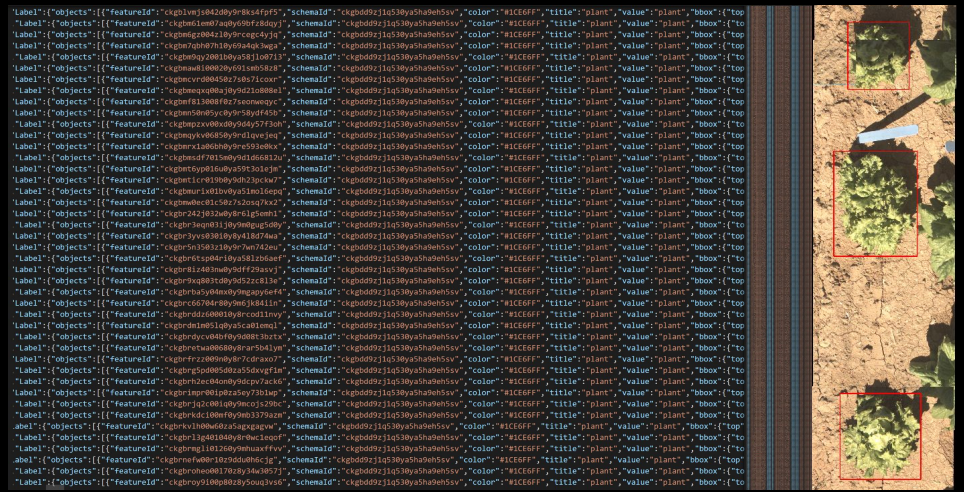




Data Preparation: Garbage in, garbage out

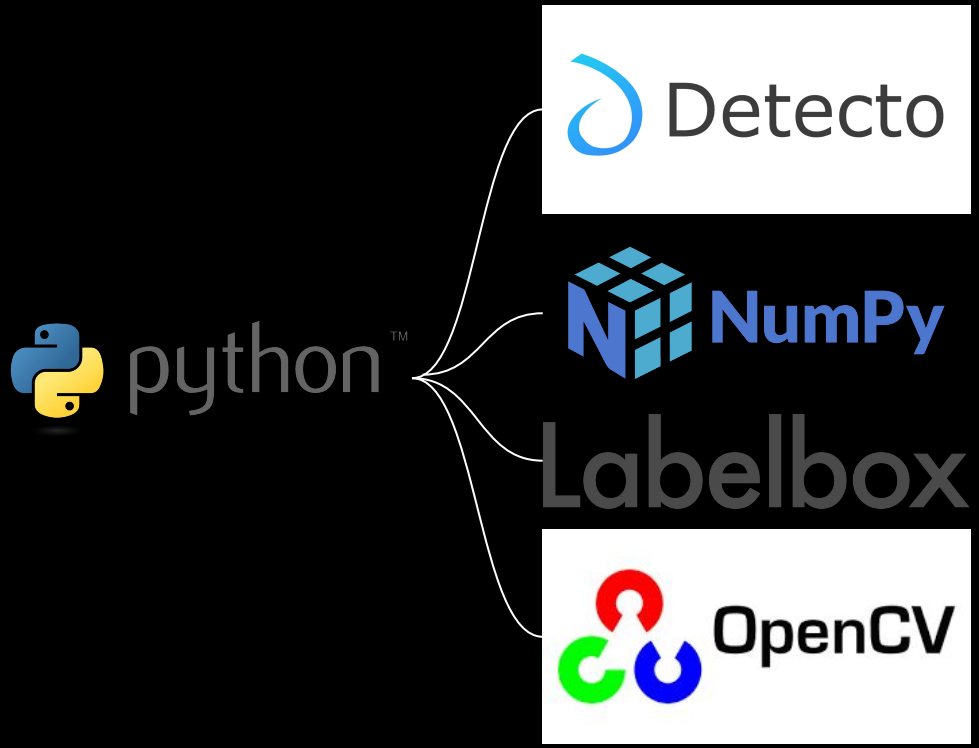
In groups of 4-5:

- Export labels
- Convert JSON file to XML
- Visualize labels for quality control



Model training: The relatively easy part

Launch Notebook: [Google Collab]



Model training: The relatively easy part

Labeling

- Labelbox

Machine learning

- Detecto

Array/image manipulation

- NumPy
- OpenCV