



Data Preparation for Machine Learning

Improve your ML models by curating your vision data

SBB CFF FFS



AI Retailer
Systems

FORTUNE
500

arm

curbFlow

Frontify®

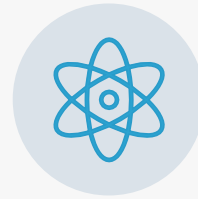
lightly.ai

Challenges with large datasets



High Costs

Up to €100k for an average 3D point cloud dataset with 10,000 samples



Overfitting of ML models

Data redundancy reduces accuracy and model generalization



Labeling time of 6 weeks

Lack of early insights from raw data affects the whole project



Redundant data **lowers** the quality of ML based-products, **cause high costs** and **slow down** product development

We filter training data from redundant data

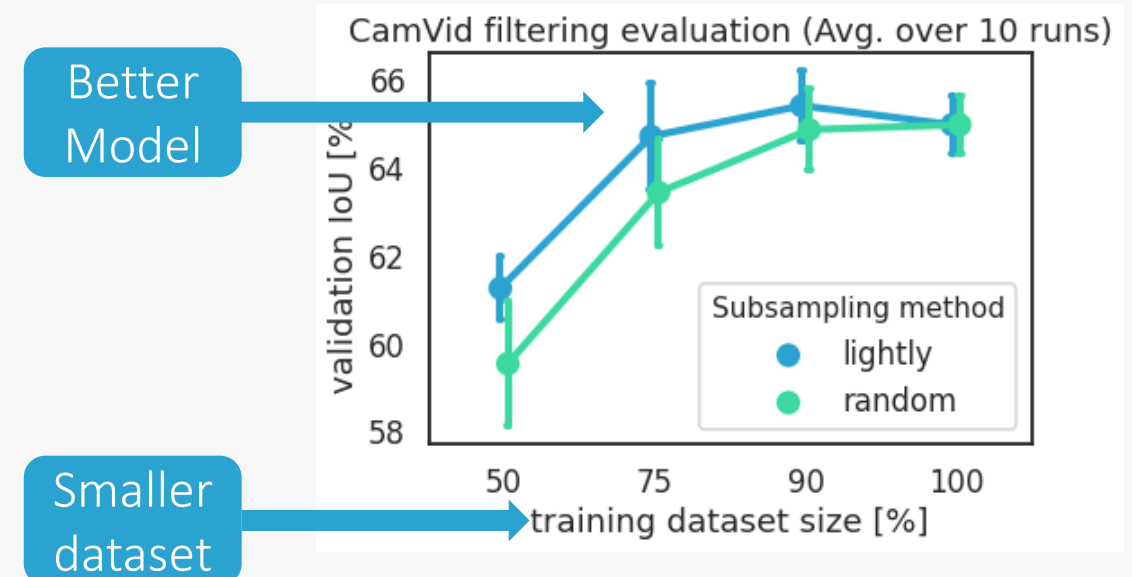
What our solution does

- Our software uses self-supervised learning
- It learns for example which images of dogs are similar and therefore less important

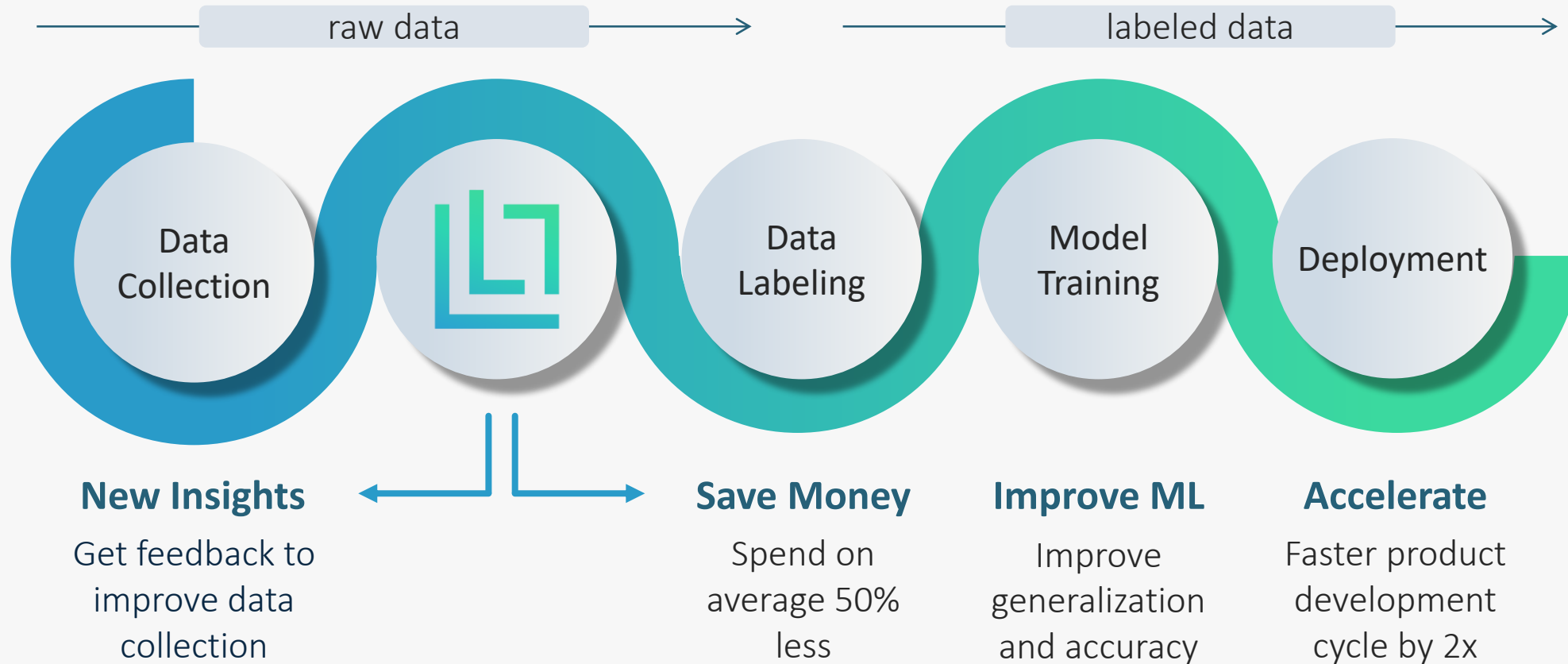


As a result our solution helps

- Get better ML models through less overfitting
- Decrease dataset size



Creating value on every value chain step



Who do we address?



Our product is especially loved by

- 1 Machine Learning/Deep Learning Engineers
- 2 Perception/Computer Vision Engineers
- 3 Data Scientists

We can help you achieve better results for the following data types



Pictures (RGB)



LiDAR



Audio



Text

Example areas of applications



Autonomous Vehicles & Robotics



Predictive Maintenance/
Visual Inspection



Marketing &
Marketplaces



Medical Image
Analysis/Diagnostics

Example
Companies



We offer smart solutions to work with data



Data Selection



Data Pool



Analytics



Get in Contact with us



Matthias

MSc. HEC Paris | B.A. HSG, VUS Harvard

LinkedIn: [/in/matthiasheller/](#)

E-mail: matthias@whattolabel.com



Igor Susmelj

MSc. ETH Zurich | BSc. ETH Zurich

LinkedIn: [/in/igorsusmelj/](#)

E-mail: igor@whattolabel.com

Build better ML based-products, accelerate your product development, and save up to 50% of your data related costs