Data Preparation for Machine Learning

Helping engineers filter and analyze training data for the best results

SBB CFF FFS

AI Retailer Systems

FORTUNE 500

curbFlow

Frontify

whattolabel.com

Challenges with large datasets

WhatToLabel

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

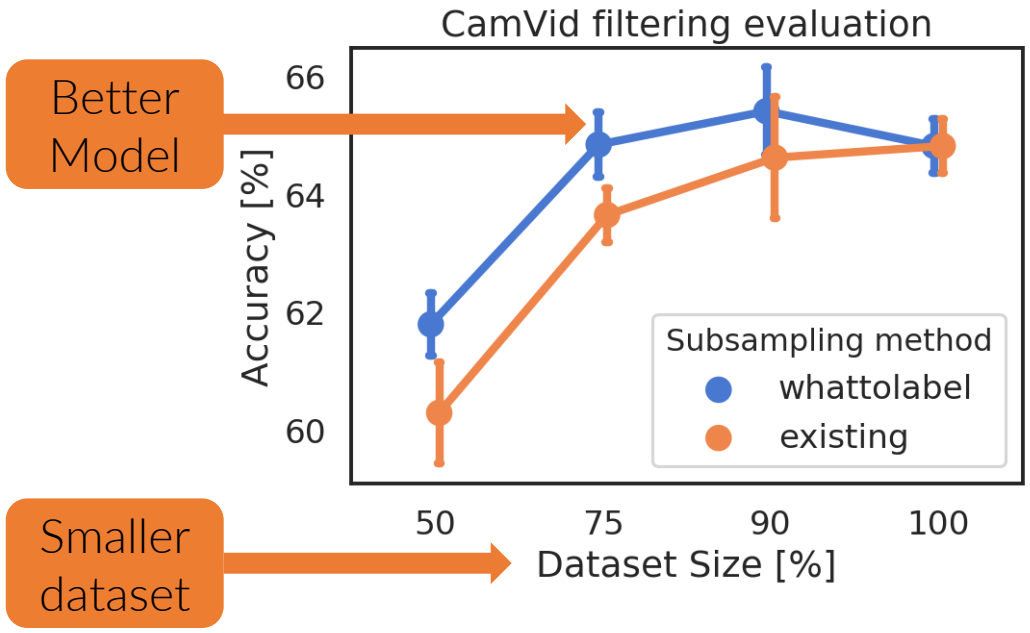
WhatToLabel

What our solution does

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

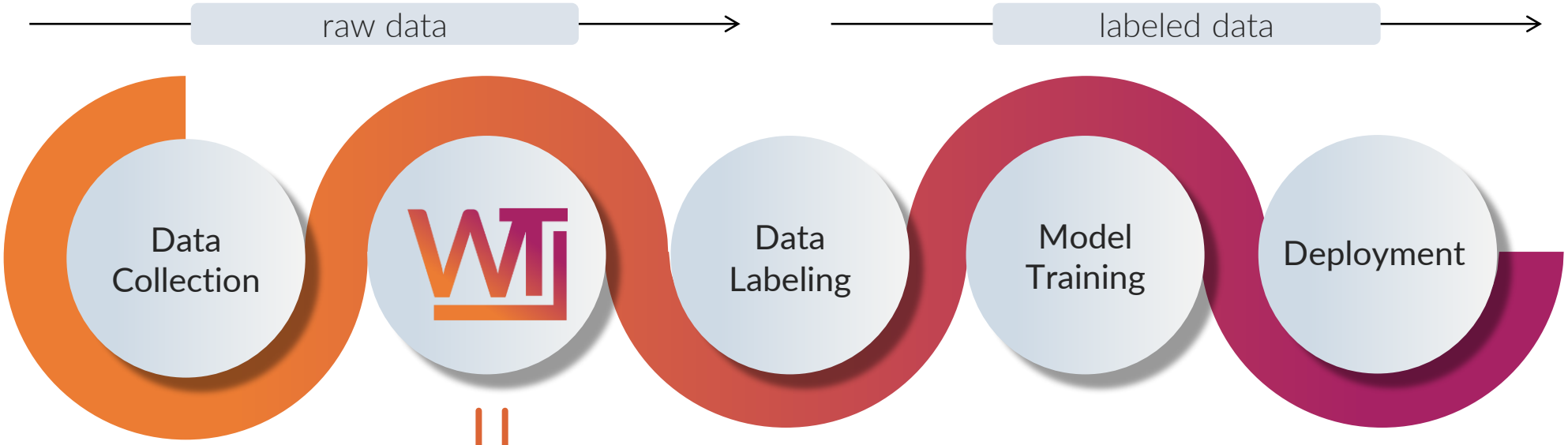
As a result our solution helps

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



Creating value on every step of the value chain

WhatToLabel



New Insights

Get feedback to improve data collection

Save Money

Spend on average 50% less

Improve ML

Improve generalization and accuracy

Accelerate

Faster product development cycle by 2x

Who do we address?

WhatToLabel

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

BMW, SIEMENS, DJI, BOSCH Invented for life, SWISS POST, UPS, SBB CFF FFS, airbnb, Pinterest, Roche, BAYER, NOVARTIS

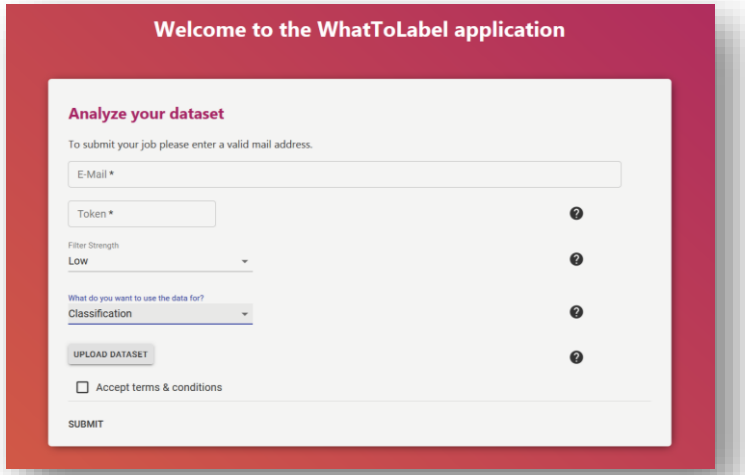
Our Product

WhatToLabel

Our AI BORIS:

Continuously learns, keeps track of what's already in the company data pool, and gets better over time

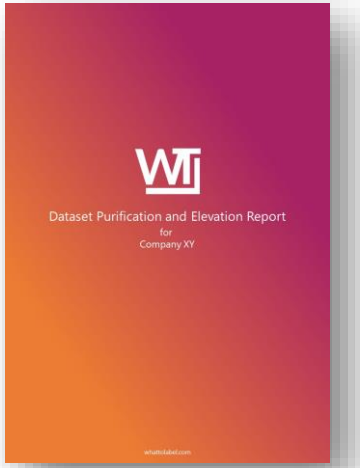
Scalable from On-Premise to Cloud

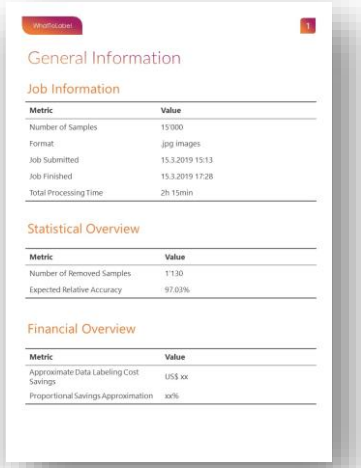


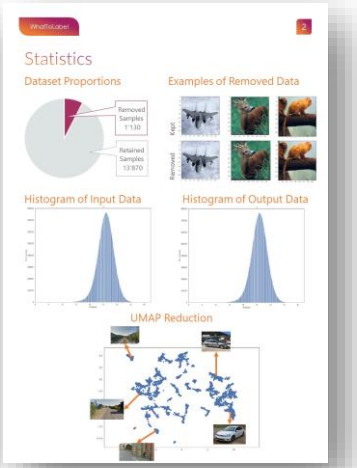
docker

kubernetes

Provides Data Analytics







Get in Contact with us

Matthias Heller

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