# Improving the production process

Dr. Franziska Horn

Motivation

Situation / Problem / Goal:

Wrong process parameters lead to the production of a lot of scrap parts. We aim to automatically detect defective products, better understand why the produced products become faulty, and automatically determine better production settings to reduce the amount of scrap produced.

85%

15%
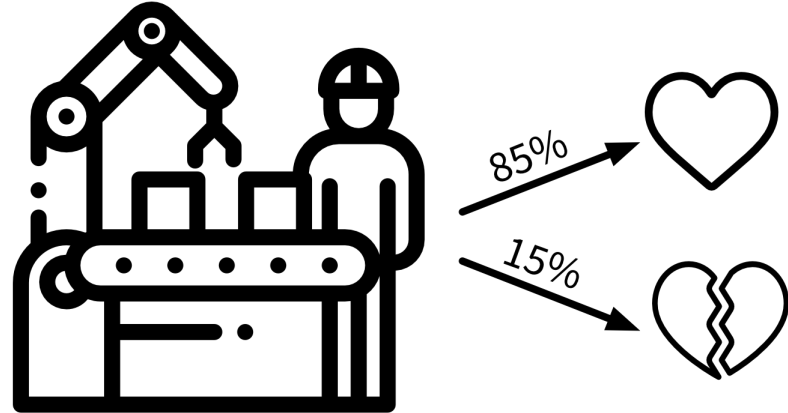
Business KPI:

percentage of scrap parts produced

Value Generation:

| X | process improvement (reduce costs) |
| | new product / feature / service (increase revenue) |

Status Quo: 15% scrap

Solution Outline

Deliverables: [X] **insights:** better understanding of root causes of scrap production   [X] Build

[X] **software:** model that predicts if settings might result in scrap (& recommends better settings)   [ ] Buy

Inputs:

[X] **(numeric) values:** raw materials + settings

[ ] image

[ ] text

[ ] other: (e.g., audio, video)

1 Data Point:
a produced product

Workflow Integration:

interact with process operating terminal to receive inputs and display prediction + recommendation for production settings

ML Solution & Output:

[ ] Dimensionality Reduction: **2D coordinates**

[ ] Outlier Detection: **anomaly score**

[ ] Clustering: **cluster index**

[ ] Regression: **continuous value:** _____

[X] Classification: **discrete value (e.g., yes/no):** product ok / scrap

[ ] Recommender Systems/Information Retrieval: **ranking of items**

[ ] Deep Learning: **other (e.g., image, text, …):** _____

Additional Steps?

[X] Explain predictions (e.g., to identify root causes)

[X] Use model in optimization (to find optimal inputs)

Challenges & Risks (+ Mitigation Strategies)

What might go wrong?	Probability:	Can you do anything about this?
might not have the right inputs to solve the problem with ML	medium	ask domain experts if they have an idea what could be important & possibly collect additional data
the available 10k data points contain many duplicates and comparatively little scrap parts (i.e., unequal class distribution)	high	systematically experiment with production settings that might result in scrap to get more diverse data; evaluate models with balanced accuracy metric
automated retraining of models will be difficult as data needs to be collected from individual machines manually	high	build a better data infrastructure where data from the machines is sent automatically to a central database
operator might get frustrated if wrong recommendations disrupt his workflow and require many extra clicks	medium	start with subtle warnings and suggestions and track how often they are ignored