DB Cargo

Eisenbahn-Bundesamt

Deep Learning for business automation

A digitization story

Christopher Wetekamp, Dr. Jan Werth

18.10.2024

eraneos

Who are we?

Your Speakers for Today

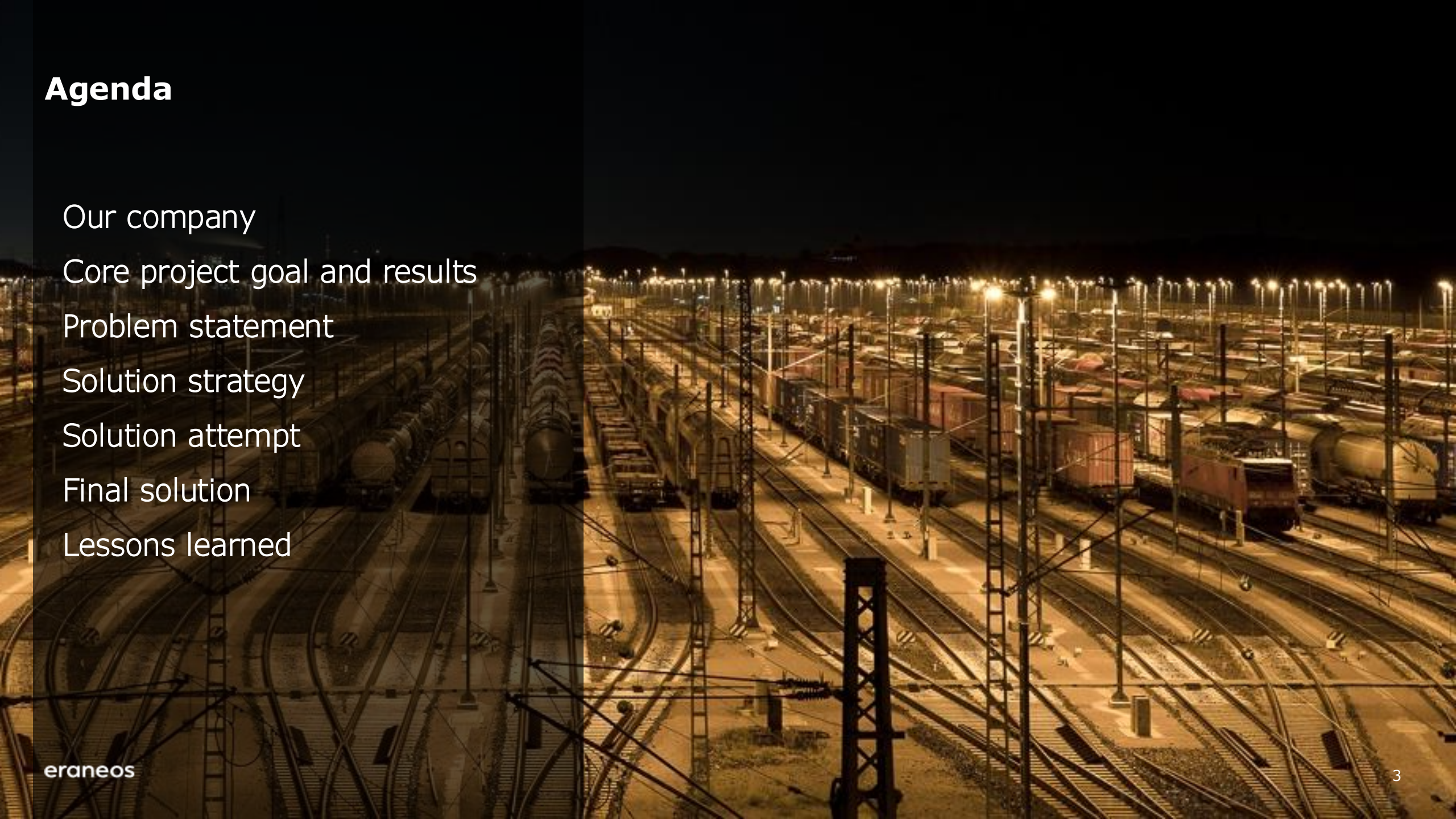
Christopher Wetekamp

Data Scientist

Dr. Jan Werth

Lead Data Scientist

Agenda



- Our company
- Core project goal and results
- Problem statement
- Solution strategy
- Solution attempt
- Final solution
- Lessons learned

We at Eraneos help our customers unlock the full potential of digital. Business and technology have been in our DNA for more than 30 years

14 offices in nine countries

1200+ dedicated professionals

2021 top consultancies join forces

150+
Data & AI Experts

500+
Data & AI Projects Globally

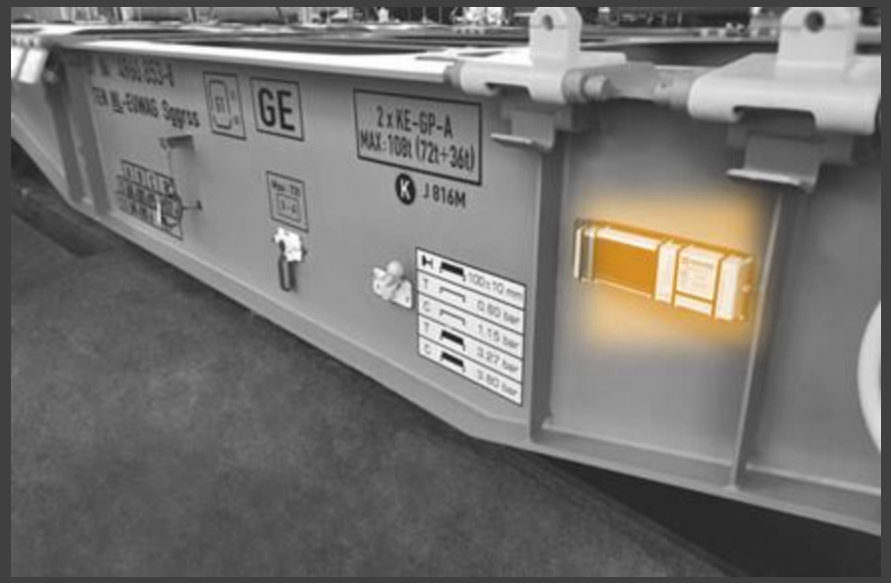
150+
Customers cross-industry

The order process at Deutsche Bahn Cargo was done completely manual | In Germany, the system was somehow running - outside of Germany, no Information was available

- No Information about wagons outside Germany
- Information exchange via mail and phone
- Manually entries from personal at the ground for business events (order completed)
- Huge gap between the events and their system entries
- Legacy System from the 1970th with over 1000 non-documented interfaces

Eisenbahn-Bundesamt

Digitization of the fleet with telemetry boxes - 60k+ wagons are equipped



60k equipped DBC cars

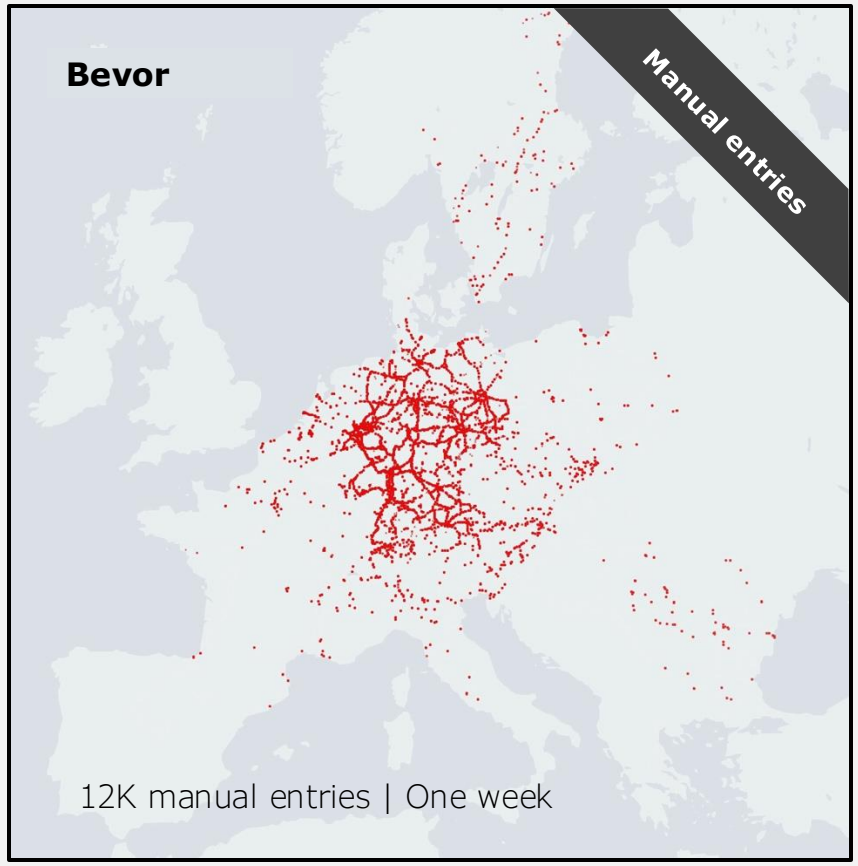
Transmission intervals are
- 10min while moving,
- 24h while parking

- **GPS signal with an accuracy of a few meters up to twenty meters**
- No track-accurate location possible
- Buildings shield the GPS signal
- However, sufficient quality for the desired goal
- Accelerometer available, but not suitable for our purposes
- Additional sensors are installed that we are not currently looking at

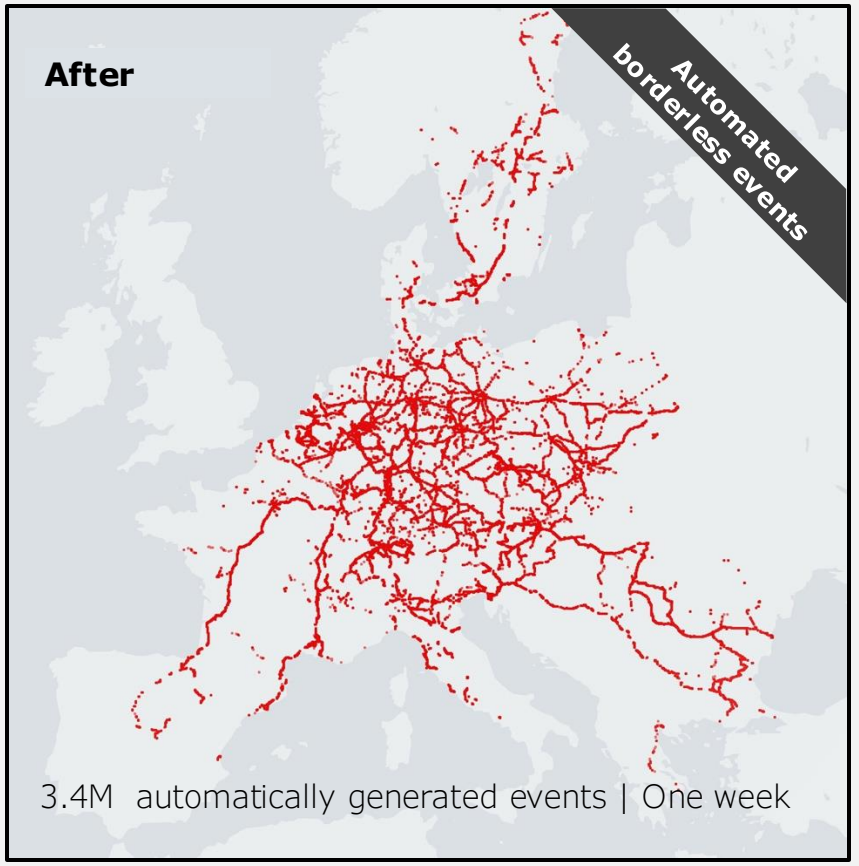
Codename VABE

eraneos

Data and information density has improved significantly as a result of VABEs



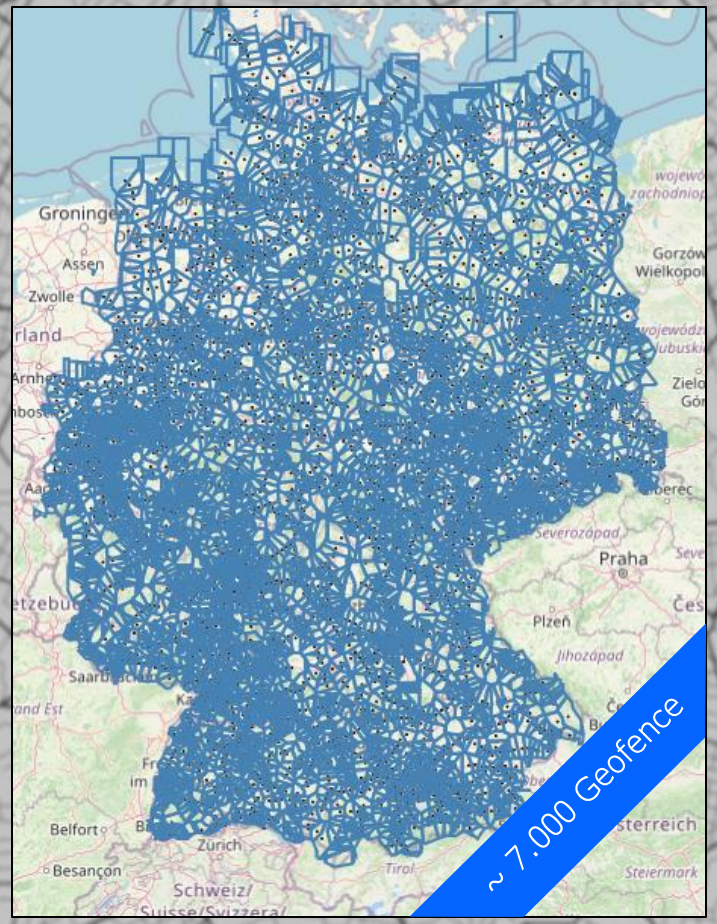
Effort
Value
Actuality
(hours to days)



Effort
Value
Actuality
(max. 30 minutes)

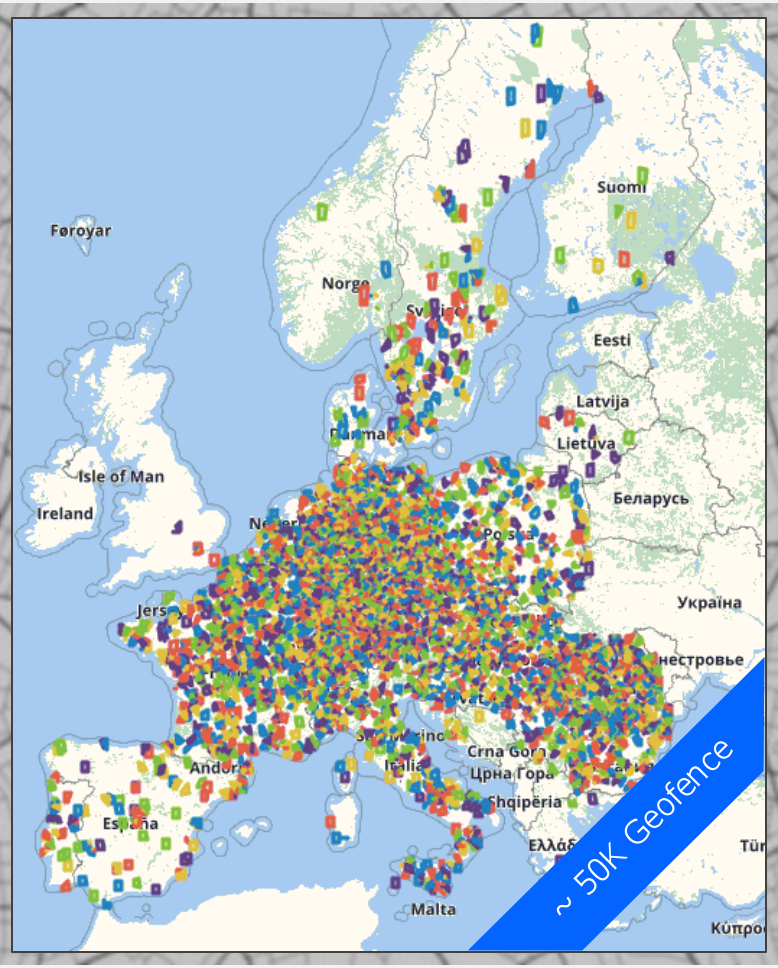
Through VABEs, we have an information gain of over 80% in order tracking throughout Europe.
The message is sent in near real-time and thus enables all parties involved to accurately track the traffic

With the geofence-mapping of Europe, we can easily match a telemetric signal to a station - **Events are generated based on the Entry and Exit of the geofences**



- Whole Europe was segmented into Geofences
- Each Geofence is drawn around a station
- There is no intersection between Geofences
- The fence density varies a lot, with Germany being one of the most densely populated countries in Europe
- For order activation and completion, special Geofences were created

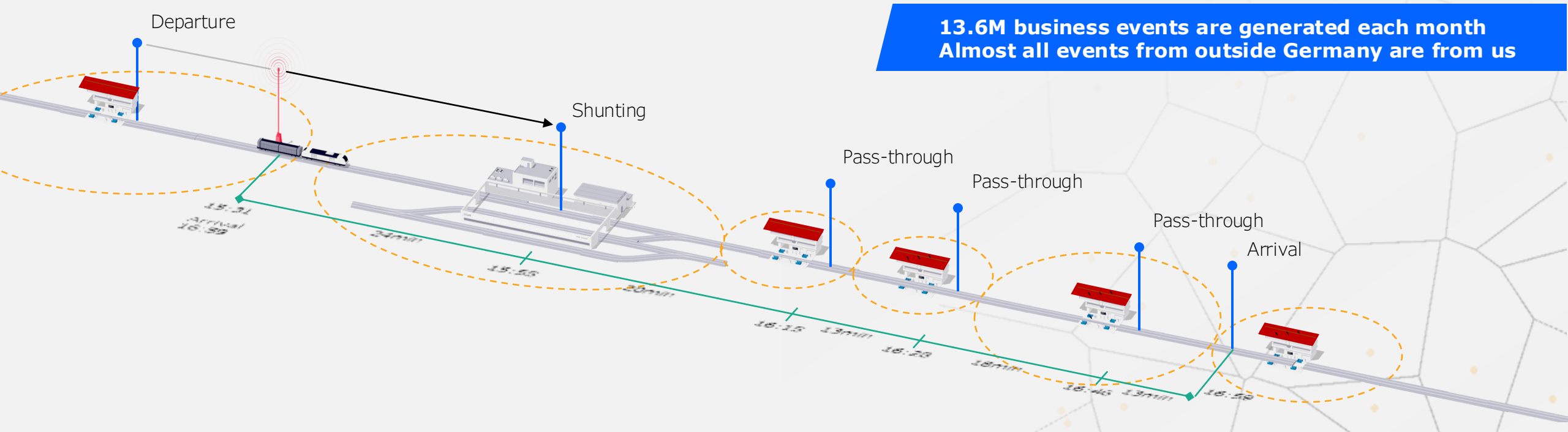
With the geofence-mapping of Europe, we can easily match a telemetric signal to a station - **Events are generated based on the Entry and Exit of the geofences**



- Whole Europe was segmented into Geofences
- Each Geofence is drawn around a station
- There is no intersection between Geofences
- The fence density varies a lot, with Germany being one of the most densely populated countries in Europe
- For order activation and completion, special Geofences were created

Using Geofences in combination with the telemetric data, **we were able to automatically generate high-precision, real-time business events for the order process**

General Geofences Order relevant Geofences	~50K (Germ. Only 7K) 1.300 ~51.3K	Departure Pass-through Arrival Border crossing	13.5M/m	Order-Activation Order Completion.	~100K/m



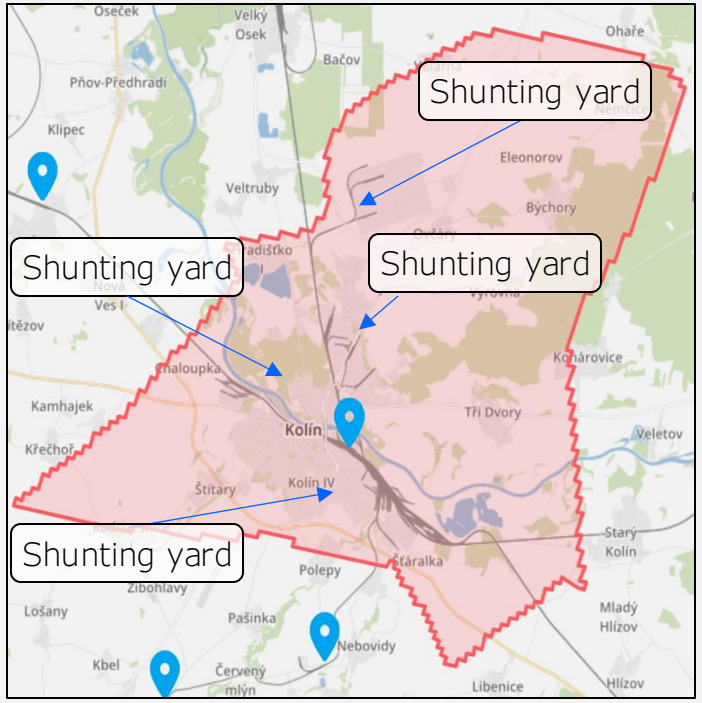
13.6M business events are generated each month
Almost all events from outside Germany are from us

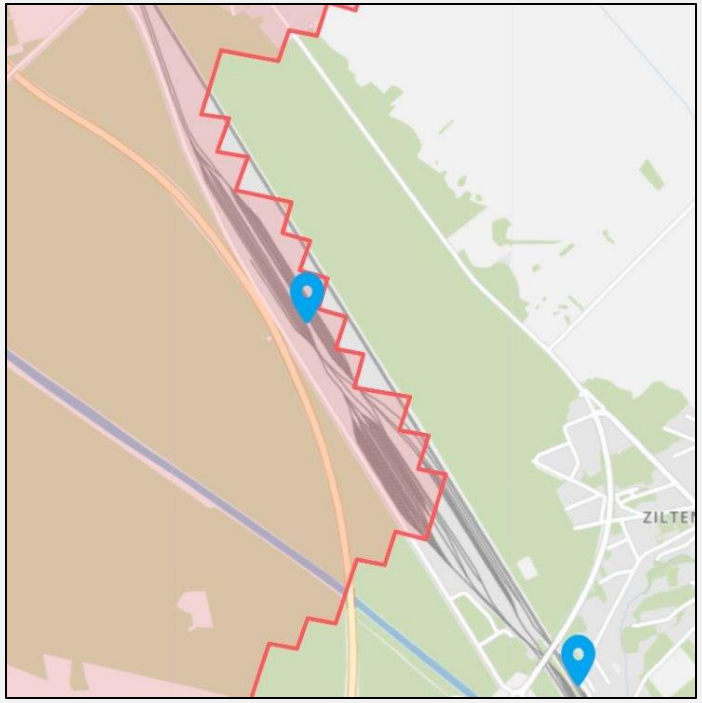
The automatic generated Geofence are not precise enough for business events connected to the order management – Expert knowledge is also not reliable

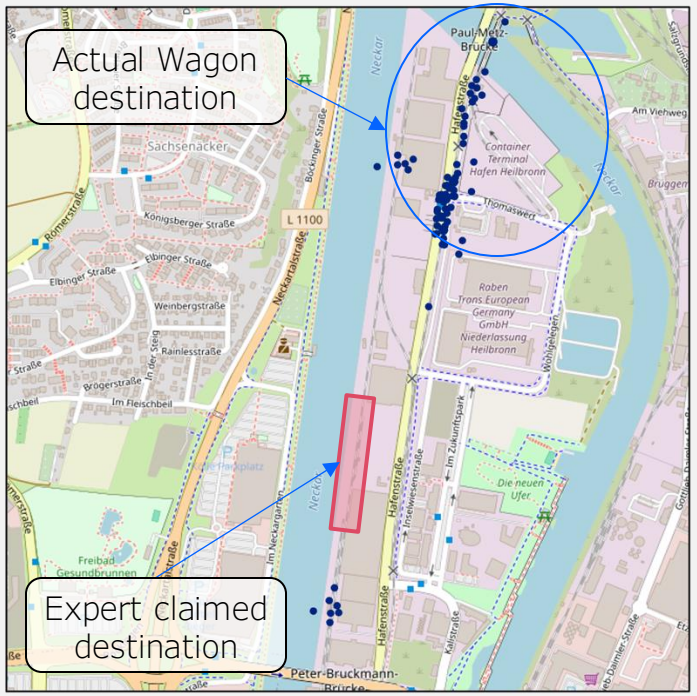
Example #1 – Large geofence
Which one is it? Within the Geofence, **multiple possible shunting yards** can be found and indicate different final destinations.

Example #2 – Unprecise geofence
The **Geofence is drawn „randomly“** while separating between the shunting yard, which could be used as a final destination, and the free track, where trains are just passing by.

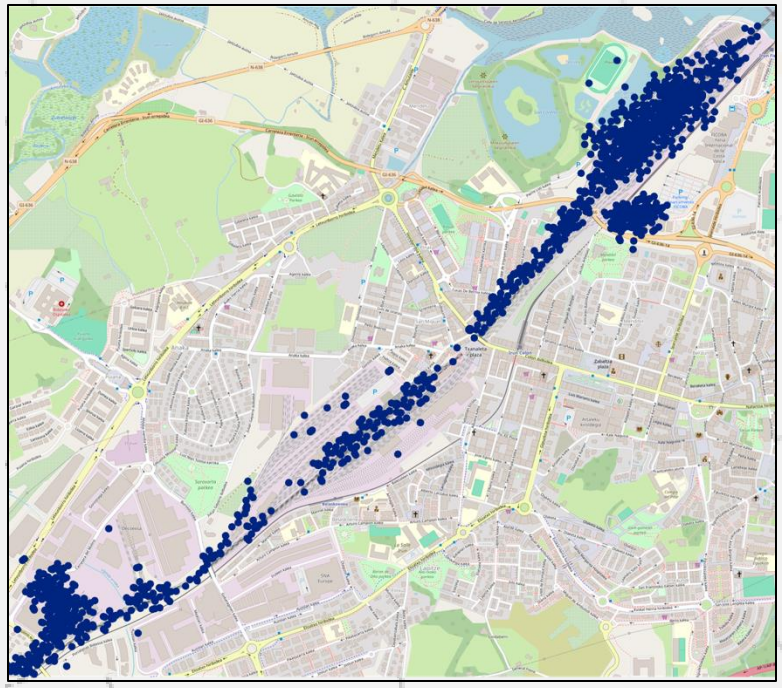
Example #3 – Expert failures
To tackle the problem, **subject matter expert** were asked to identify order relevant destinations. However, this information was not reliable.



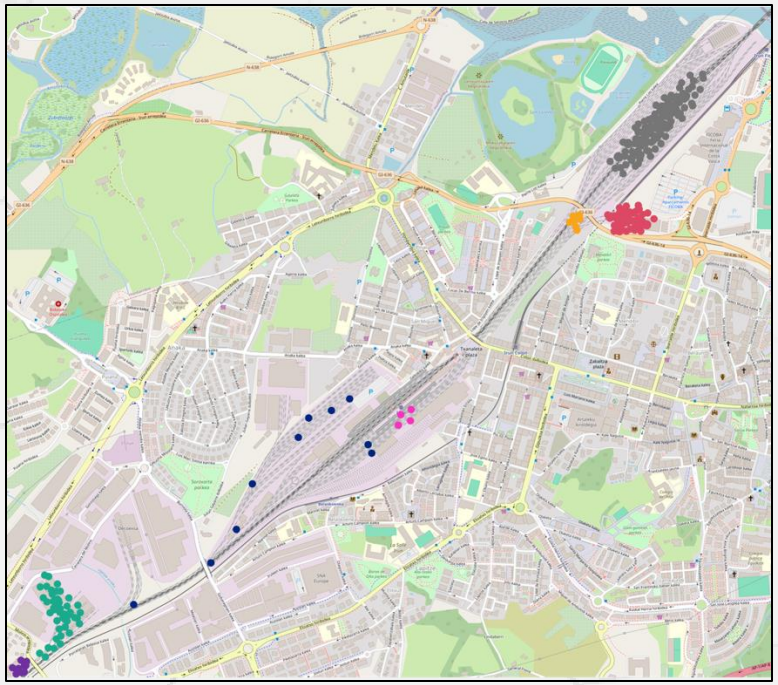




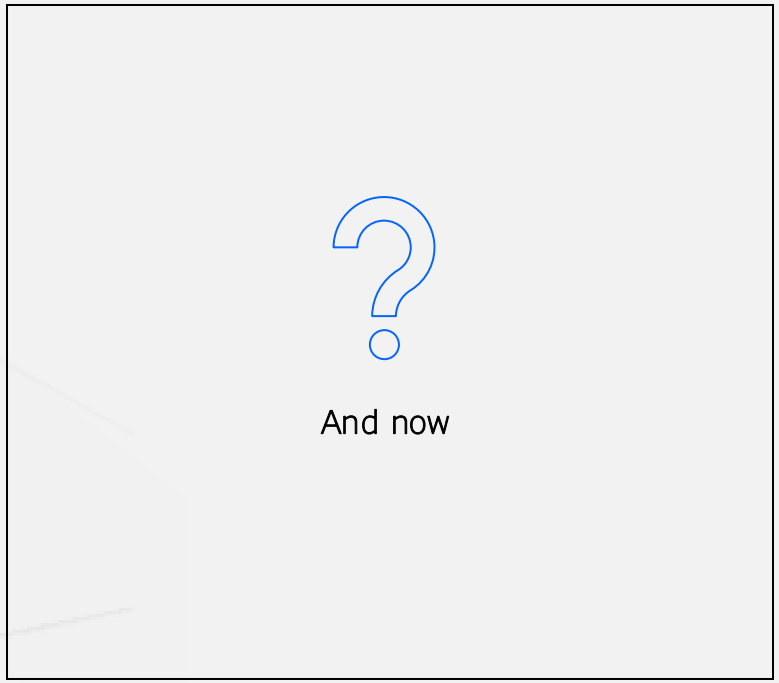
Identifying actual action locations by using multi-layered GPS clustering to determine the production locations



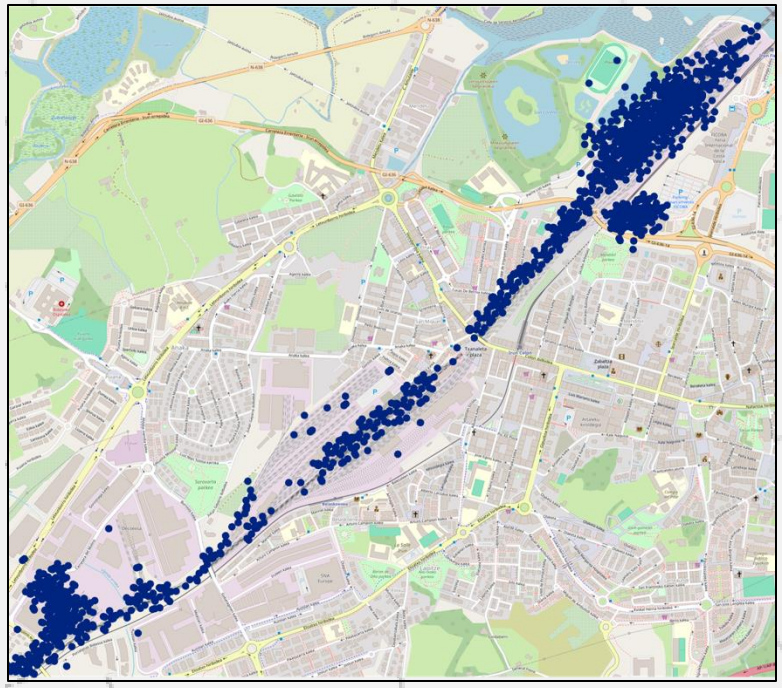
In the target area, the **GPS points** are **concentrated** in certain areas and allow a differentiated view of the target area.



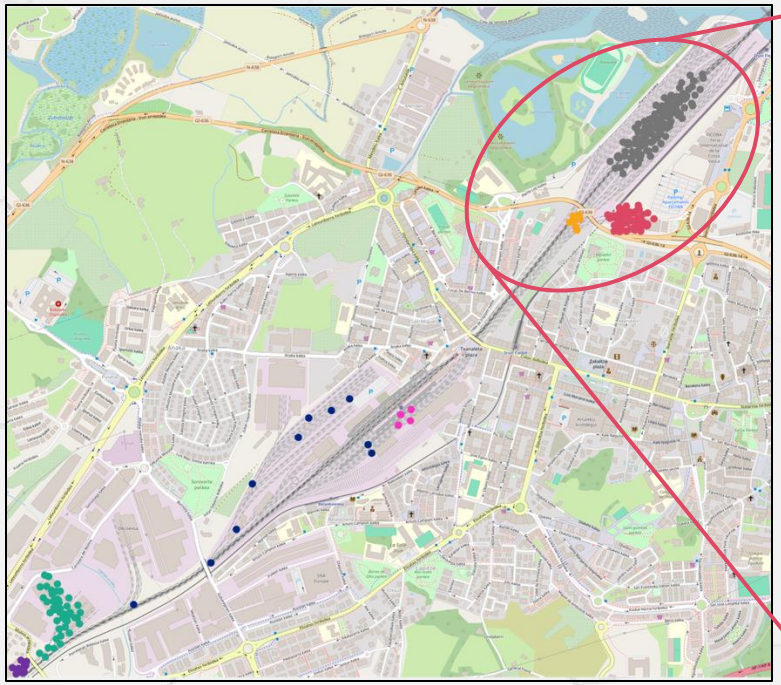
Multi-layered clustering enables the extraction of the underlying positions.



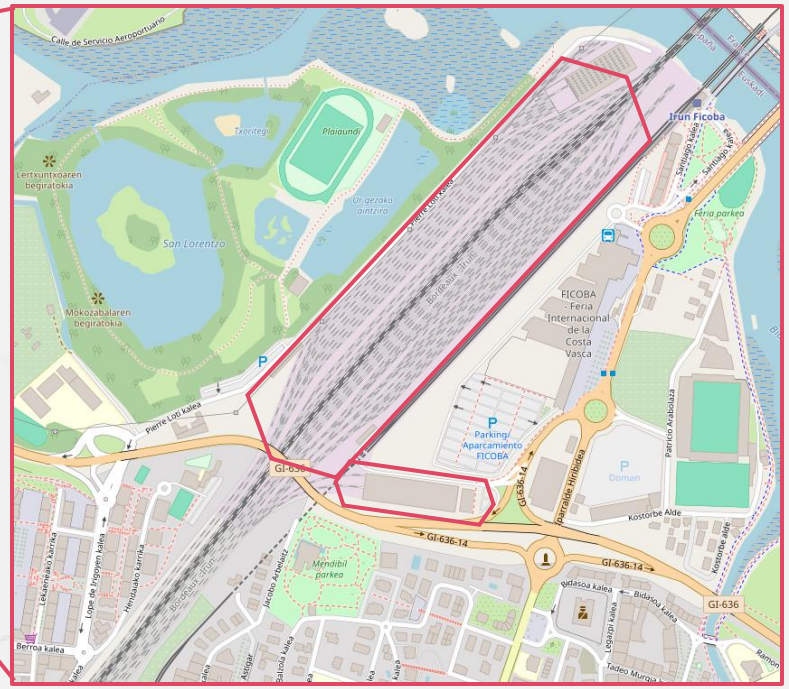
Identifying actual action locations by using multi-layered GPS clustering to determine the production locations



In the target area, the **GPS points** are **concentrated** in certain areas and allow a differentiated view of the target area.

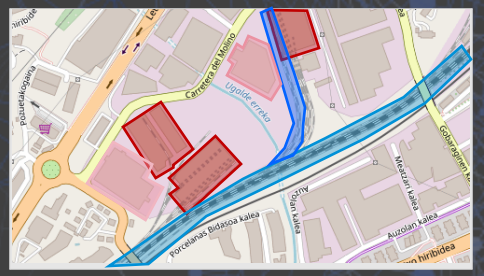
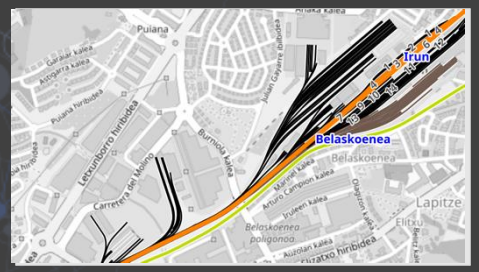
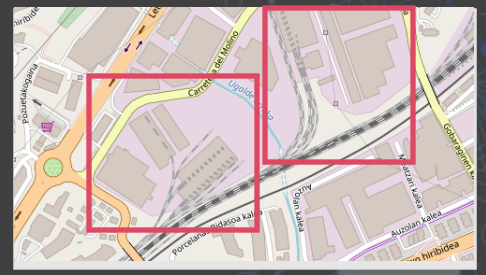


Multi-layered clustering enables the extraction of the underlying positions.



The identified clusters must be **further segmented** to determine the actual destinations.

As data scientists, we open our toolbox and unpack the cool algorithms

	Instance Segmentation Object Masks	**Graph Neural Net** Relations	**Object detection** Structures
Solutions			
Key Components	Detect and mask **railway-specific objects**. Use the **intersection** of object mask and cluster polygon to classify a cluster.	Capture the **underlying network structure** of railways. Use the **different track networks** to classify the clusters.	Locate and classify **key structures.** Use the **intersection** of object mask and cluster polygon to classify a cluster.

On the way solving the problem with a deep learning solution **we realised, that we already achieved our goal**

Clustering

We used multi-layer, feature-supported clustering to identify possible clusters for order-relevant events

The clusters in this form **could not definitely be linked with the order process**

Labelling

We knew from countlessly inspecting network maps that the cluster could be classified to be clearly linked to the order process.

We spend hours **manually labelling** prepared images to generate a training set.

Results were getting better and better

Feature Extraction

To boost the manual labeling process, we started to **enrich the data with additional features** to be able to build a labeling triage.

One Key source was the **graph information** from OpenStreetMap.

Features were number of tracks, track ends, warehouses, …

Tree Learner

Tree learner
Extracting all those features, we set up a pipeline to preannotate clusters, boosting up the manual labelling.

But lo and behold, the **prelabeling was of such high quality**, that we already achieved our goal

How well does it work?
Testing a ML System by brute force

By the books, a ML system is testes with properly labelled data and a well setup train test validate split and a powerful cross validation

However, that needs a well-documented ground truth.
Unfortunately, the ground truth was unreliable from the get-go.

The solution was the engineering way. If it works, it works.

We set up a quality measuring system, which analysed the actual order activations and completions at each location against the orders in the management pipeline.

This allowed us to identify the correct geofences, any shift in the production process at the location (e.g., shift of unloading location) and falsely matched orders

The results lead to a full European roll out

Deep learning is often viewed sceptically by the developer and software engineering community — think management

Let’s talk

Contact us for **further discussion**

Christopher Wetekamp
Senior Data Scientist

Christopher.wetekamp@eraneos.com
+49 170 9571 564

Dr. Jan Werth
Lead Data Scientist

jan.werth@eraneos.com
+49 171 6415 807