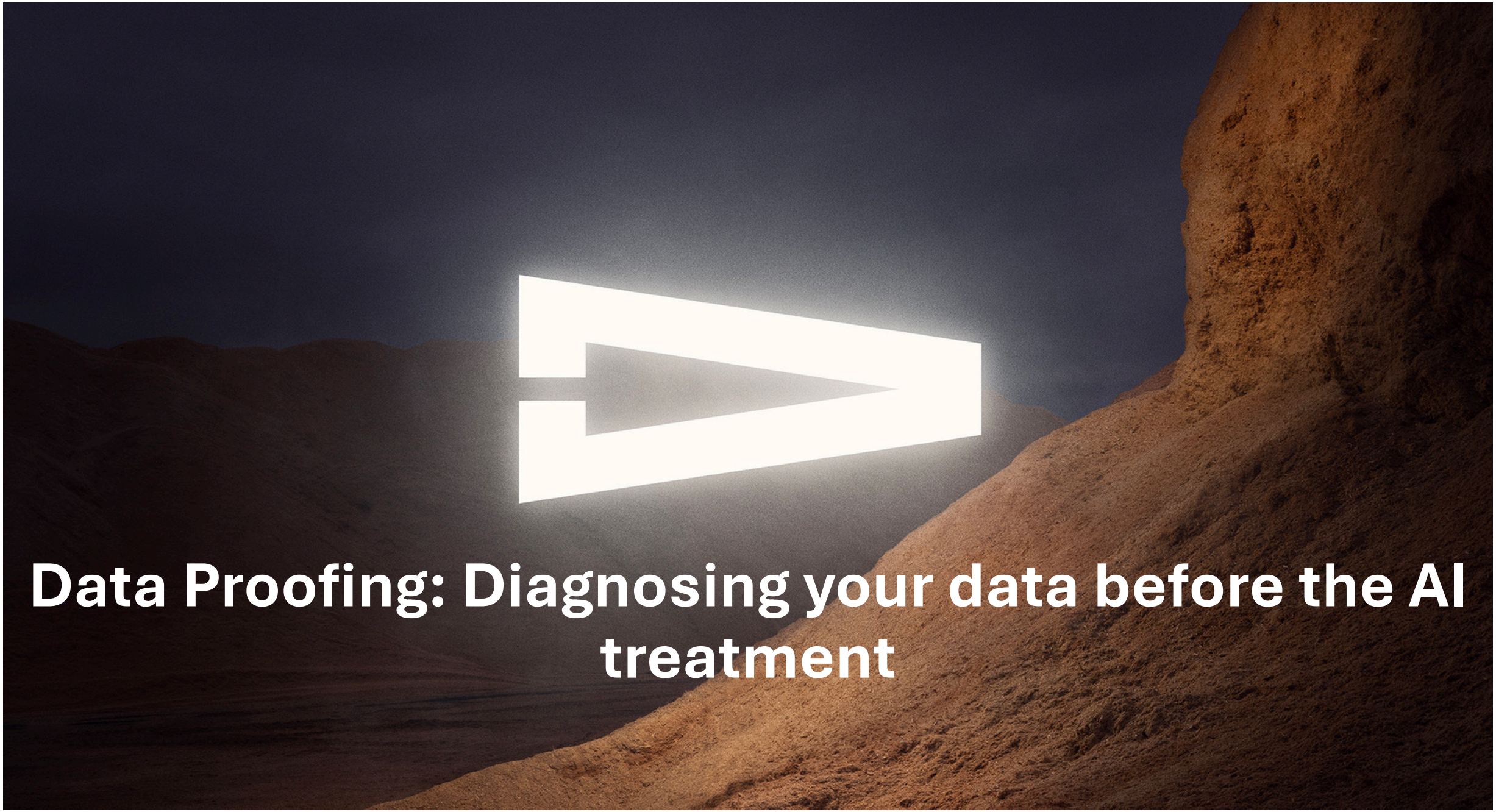
SCANDBIO


Data Proofing: Diagnosing your data before the AI treatment

Christian von Koch

Senior Machine Learning Engineer

MSc Industrial Engineering and Management with a major in AI and Machine Learning

Image segmentation, optimization, demand/sales forecasting, S&OP planning, NLP and language processing

Experience in large-scale development projects

Previously at Combient Mix / Silo AI, at Elvenite since April 2024

Elvenite

SCANDBIO

Anna Warell

Sales and Operations Planning Manager

Master in Industrial Organization and Production and
Bachelor's degree in Mechanical Engineering

Sales and Operations Planning Manager at Scandbio since 2020

Elvenite

SCANDBIO

Agenda

01 Scandbio and their challenge

02 Data Proofing

03 Conclusions & reflections

Scandbio and their challenge

Scandbio

We are a company and wants to be part of the transition to a sustainable society by developing local bio-based residues into climate-smart energy – heating pellets mainly!

We have multiple production sites and supply a wide range of customer segments with different usage patterns and different seasonality

A complex situation to forecast future needs, especially long-term forecasting – an area for possible improvement for us!?

Scandbio

We changed our ERP system to Infor M3 2023/24 together with Elvenite

We also changed our BI system to improve in many areas

Forecasting long term is one of the keys to be successful and we wanted to investigate how Machine learning could help us to improve our forecasting performance

We have together with Elvenite explored our data and evaluated the potential for ML solutions

Data Proofing

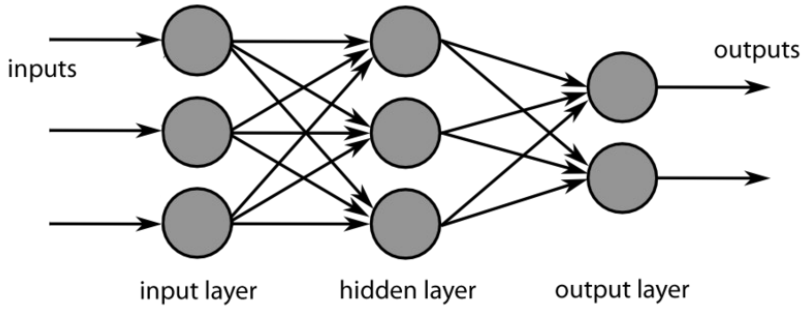
How AI models work

Data type in

Text
Video
Images
Audio
Tabular

Numbers in

1,
6,
17,
20,
54,
1025,
…,



Numbers out

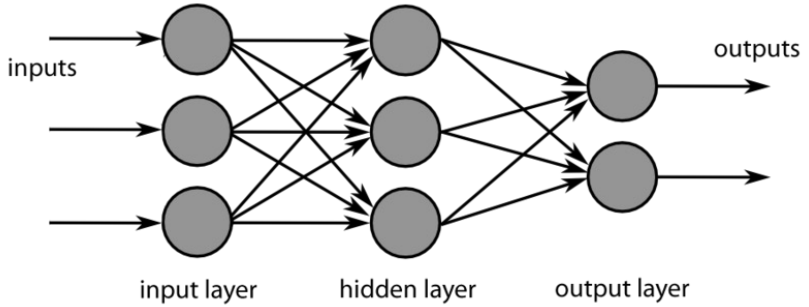
2,
9,
875,
43,
56,
87,
…

Data type out

Text
Video
Images
Audio
Tabular

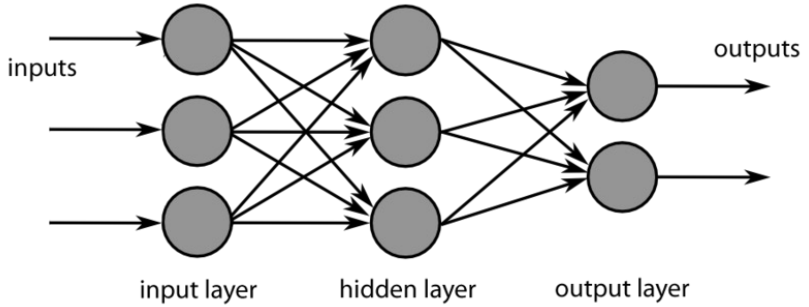
Example: Large Language Models (LLMs)

Paris is the



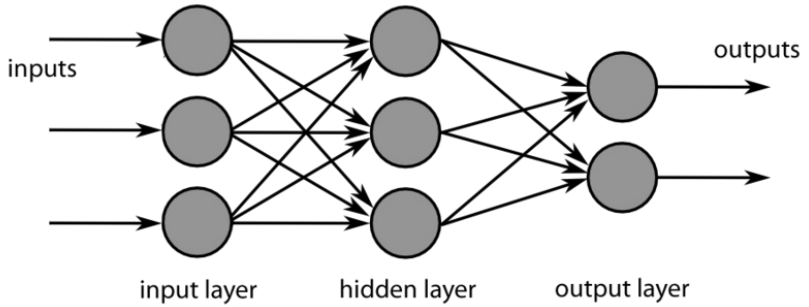
capital

Paris is the capital



of

Paris is the capital of



France.

Diagnosing your data before the AI treatment

“Doctor, I have a soar throat”

“Let’s first diagnose you and find out why”

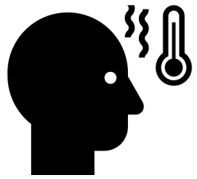
“Doctor, my forecasts are all over the place”

“Let’s first analyse your data and find out why”

Symptoms

Tests and diagnosis

Treatment



Business problems

Data Proofing

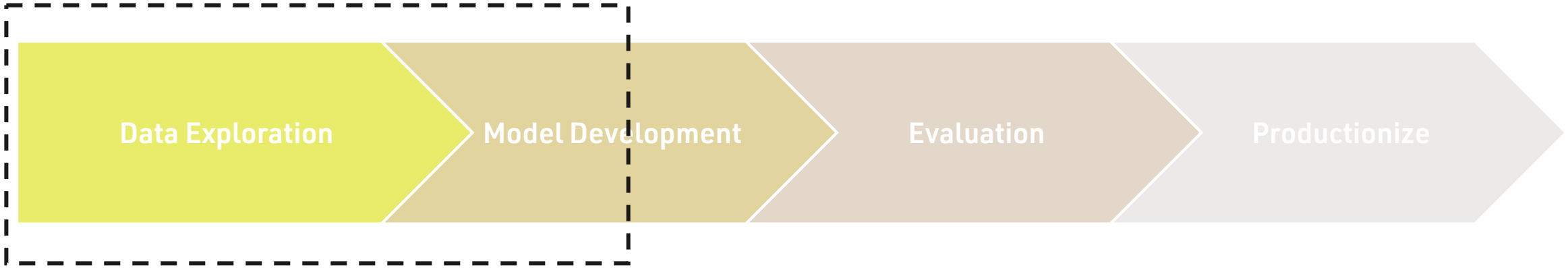
AI models

Elvenite

SCANDBIO

The AI project process

Note: These are iterative processes and usually it's not a straight line in between the phases.



***Data proofing** covers the data exploration phase and, if data is sufficient, parts of the model development phase*

Common ways to predict

- Own experience
- Same as last year
- Moving average

Risk factors

- Personal dependence
- Time-consuming
- Hard to see complex dependencies
- Already old when it’s done

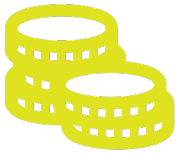
What does it cost to have inaccurate forecasts?

Excess inventory

Lost opportunities

Customer dissatisfaction

Poor operational planning



Reduced profit margins

Three data sources used in the project

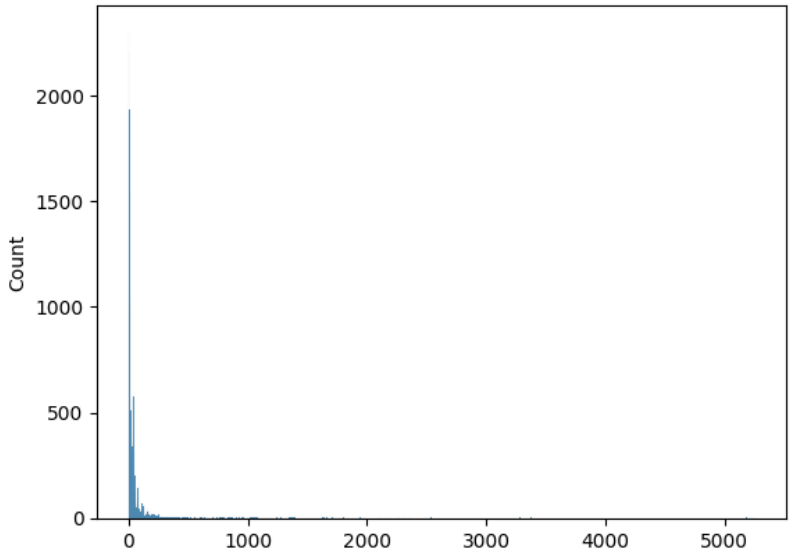
Historical sales data: Extracted from Scandbio's data warehouses originating from M3.

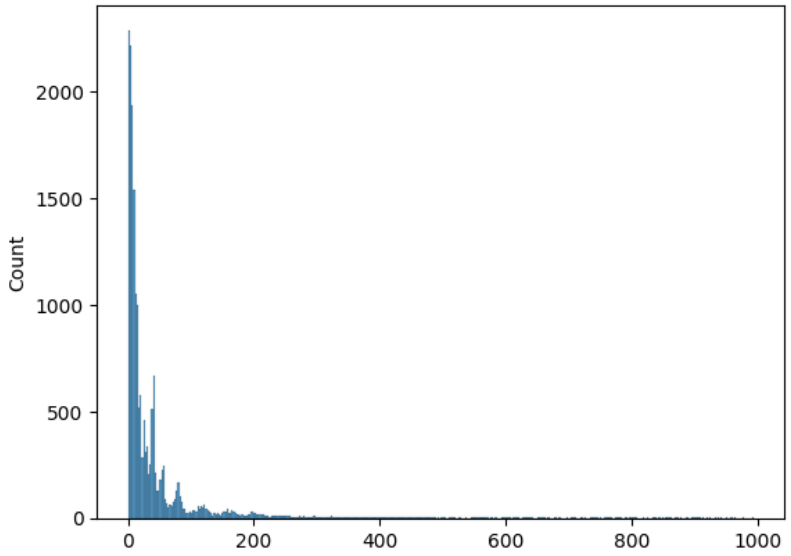
SMHI energy index: Monthly data from SMHI of the energy index describing energy needs.

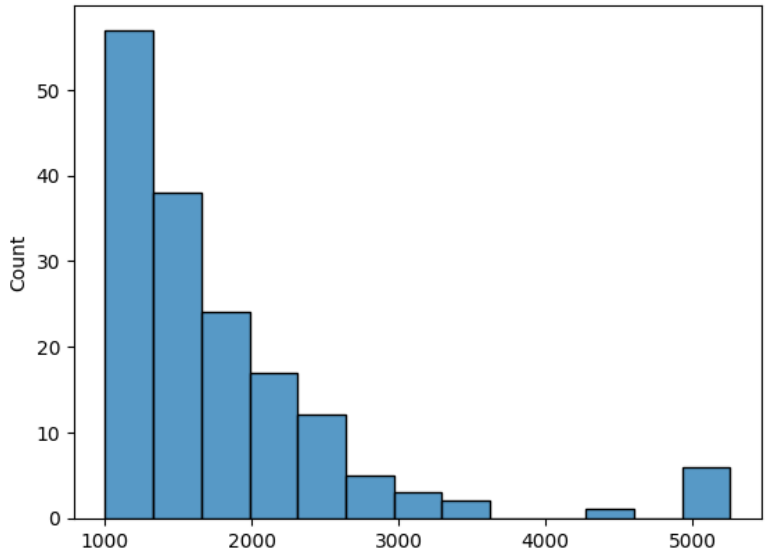
Historical forecasts: Monthly forecast documents extracted from the CRM system.

Monthly volumes sold in different ranges

Long tail distribution can cause a problem for ML models



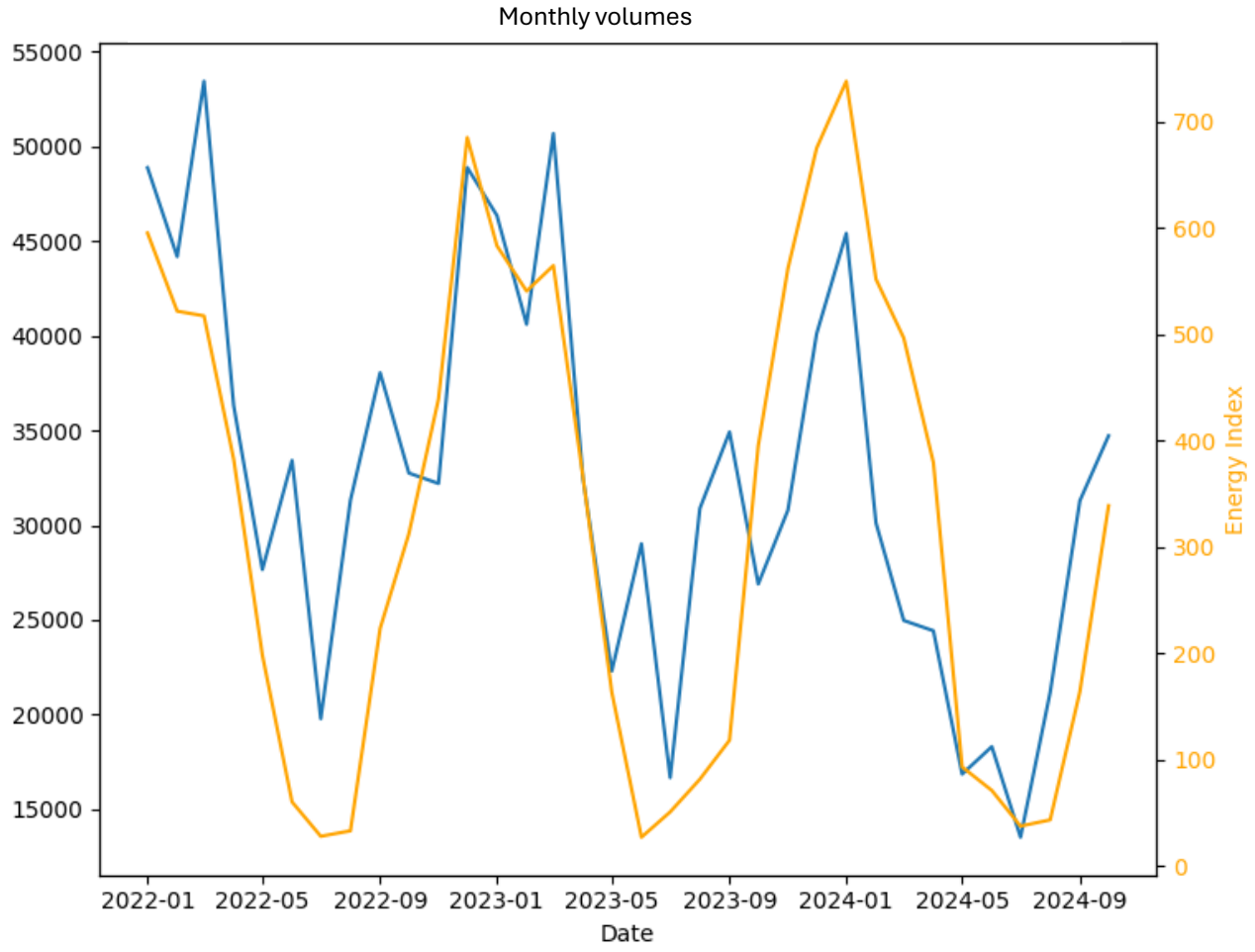




Elvenite

SCANDBIO

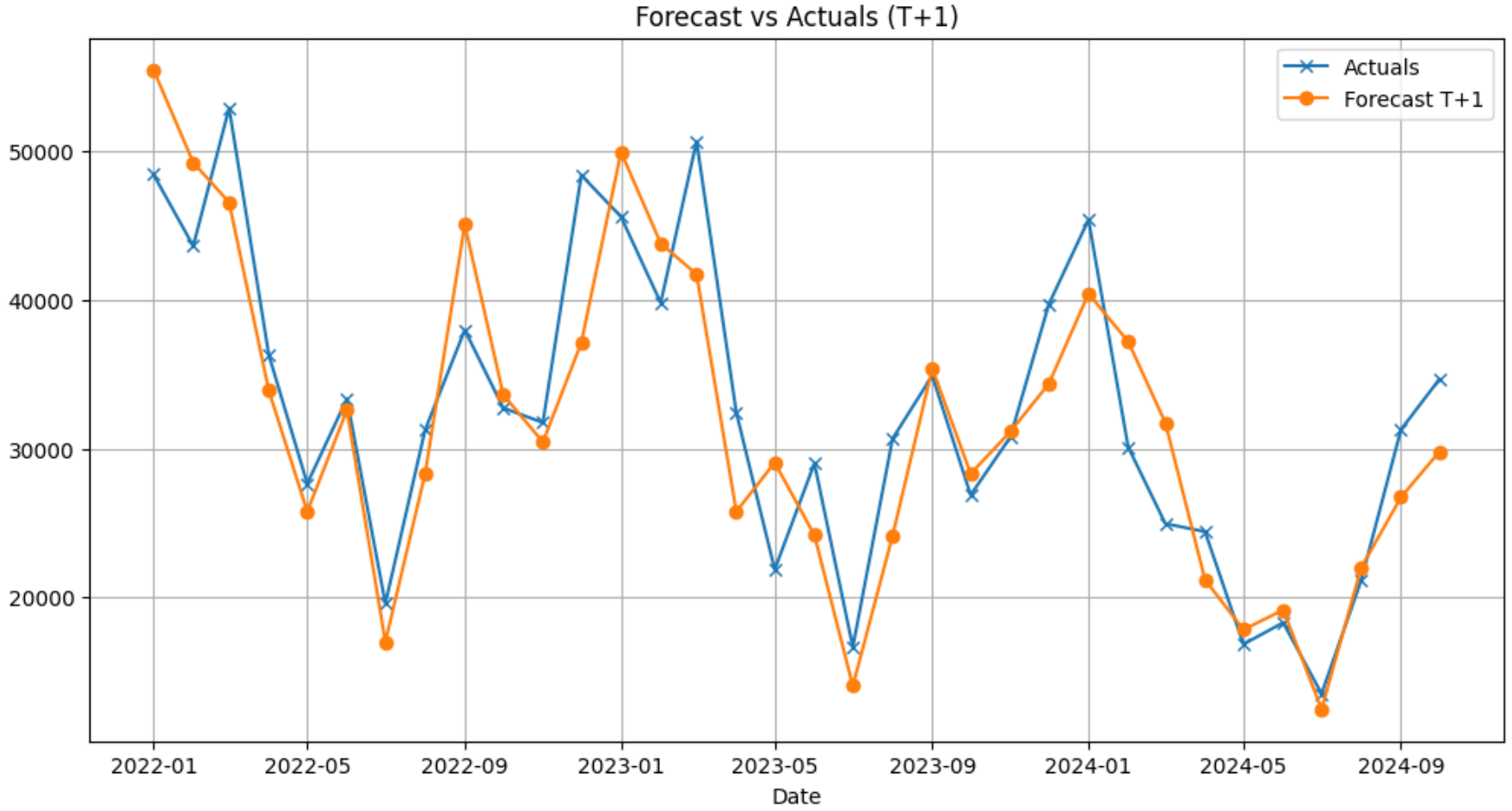
By combining datasets, several comparisons and insights can be uncovered



Elvenite

SCANDBIO

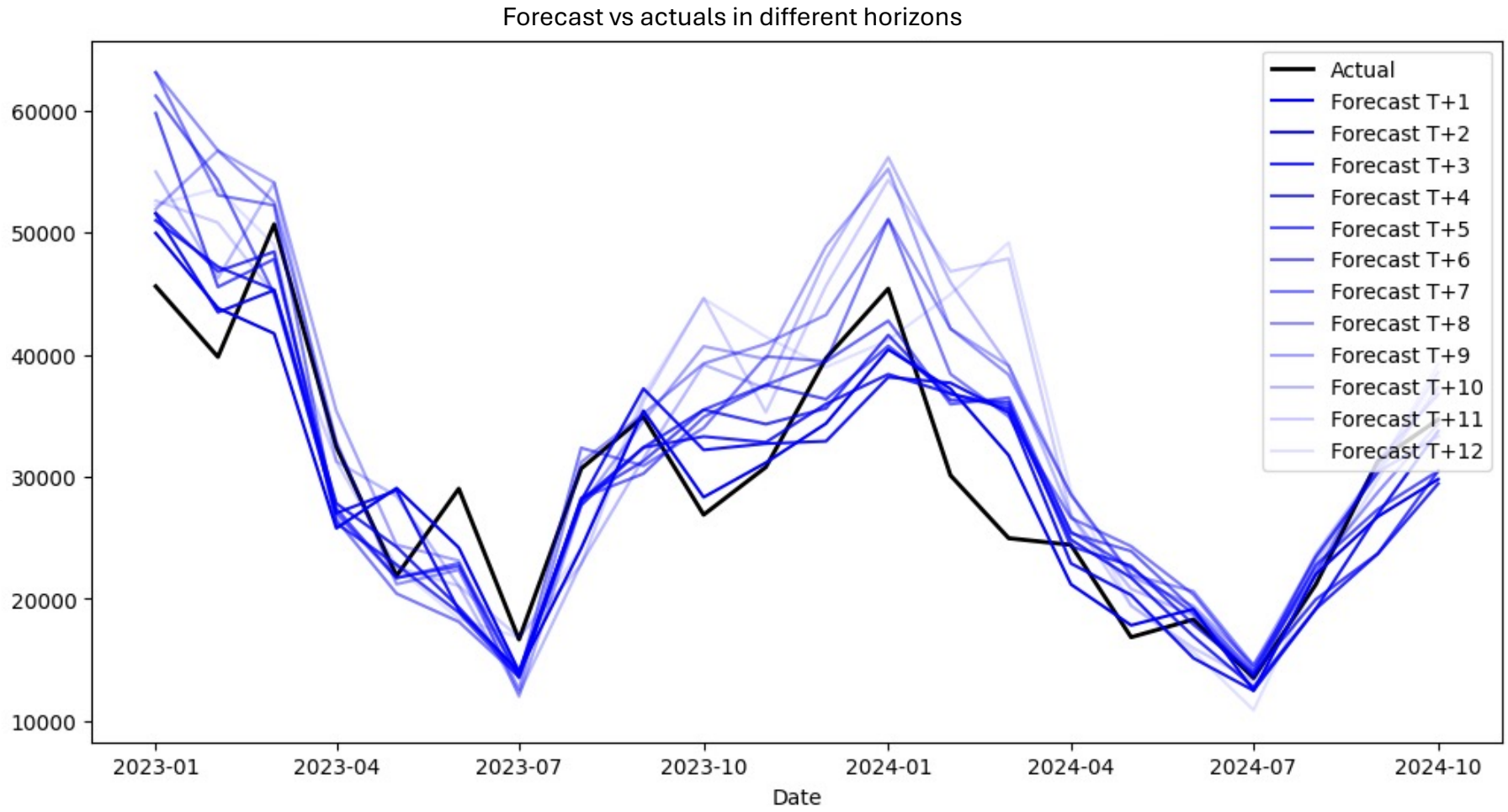
By combining datasets, several comparisons and insights can be uncovered



Elvenite

SCANDBIO

Forecasting naturally gets worse the further away in time you need to forecast

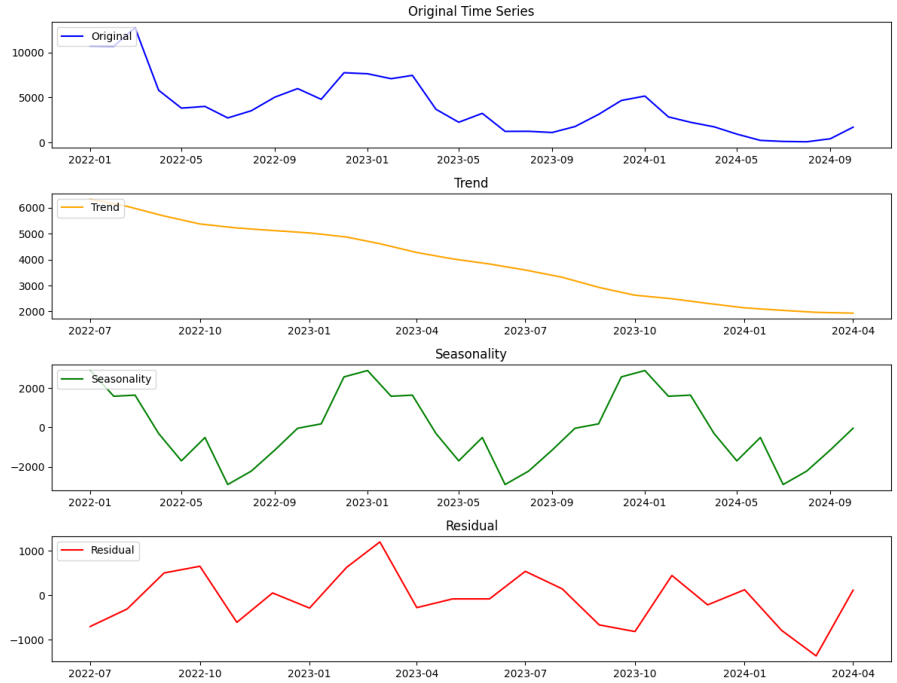


Elvenite

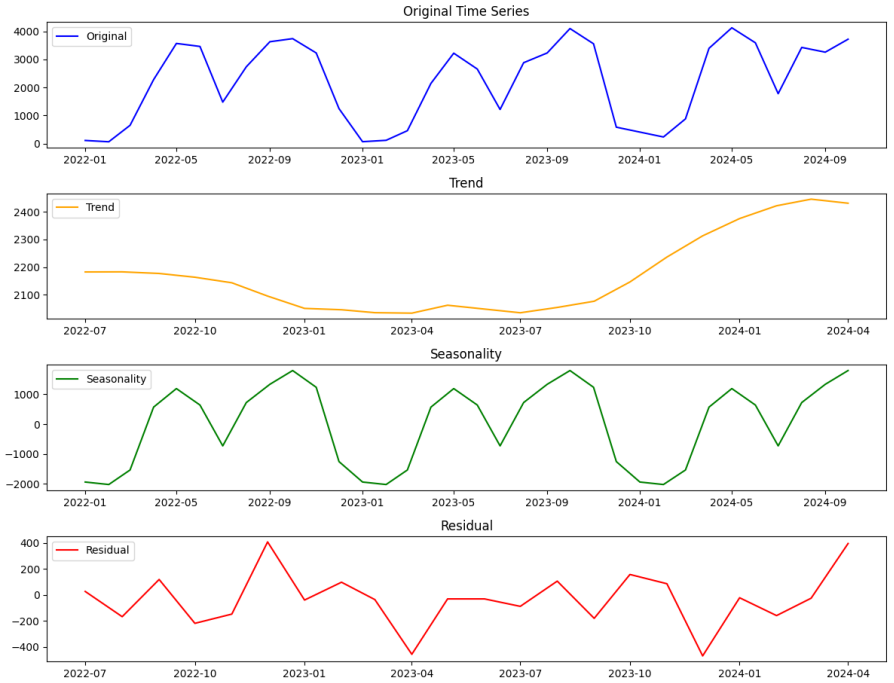
SCANDBIO

Different customer groups show different degree of seasonality and trends

- Customer group X



- Customer group Y



Elvenite

SCANDBIO

Solution? Use a machine learning model able to keep track of all the different dependencies and automize the forecasts

Sales article?

Customer group segment?

Temperature?

Sales 12m back?

Winter?

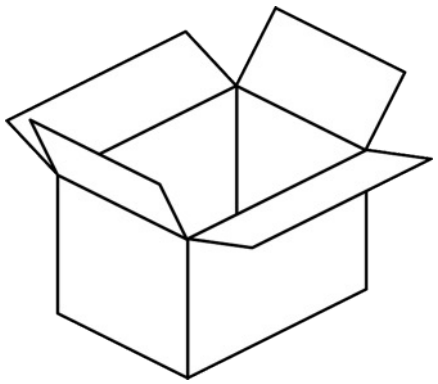
...?

Sales 3m back?

Sales article group?

Sales last month?

Holidays?



Summer?

Price?

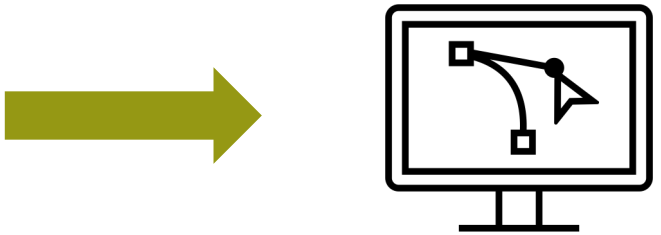
Elvenite

SCANDBIO

Supervised Learning: Usually, Machine Learning models are trained under supervision, we teach the model what to learn

Features | Label

Article	ArticleGroup	QualityGrade	Month	Quantity (kg)
Bacon	Meat	4	July	200
Minced beef	Meat	3	March	200
Ribs	Meat	2	December	5000
Ribs	Meat	5	June	50
Ham	Spread	2	September	200
Ham	Spread	4	March	200
Fresh liver pâté	Spread	1	January	5
Fresh liver pâté	Spread	4	January	150



Predicting: When the Machine Learning model has been trained, we can ask it for predictions based on the input data

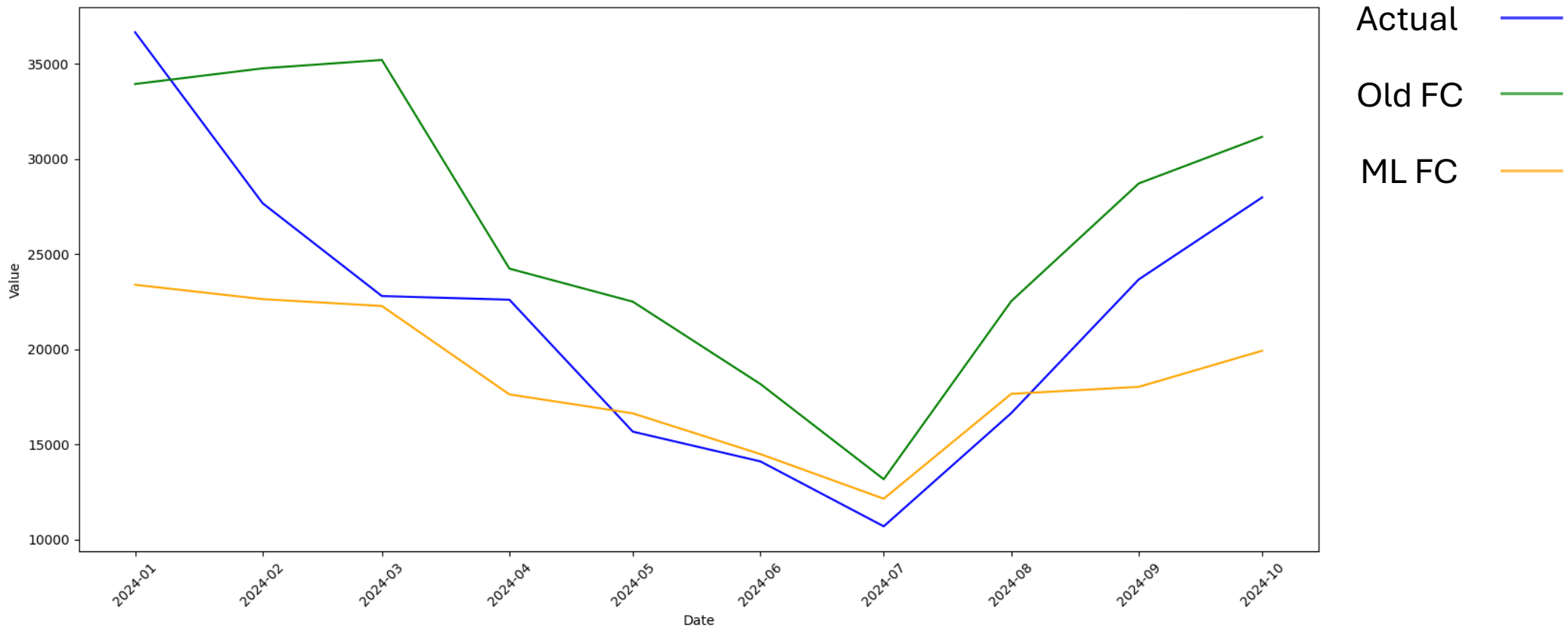
Article	ArticleGroup	QualityGrade	Month
Ribs	Meat	5	December

*"Quantity will be **400 kg** sold with an **87%** probability"*

Note here that the quantity column is not present, this is the variable we want the model to predict

T+6 horizon

The ML model showed performance gains of up to **20 %** on average monthly

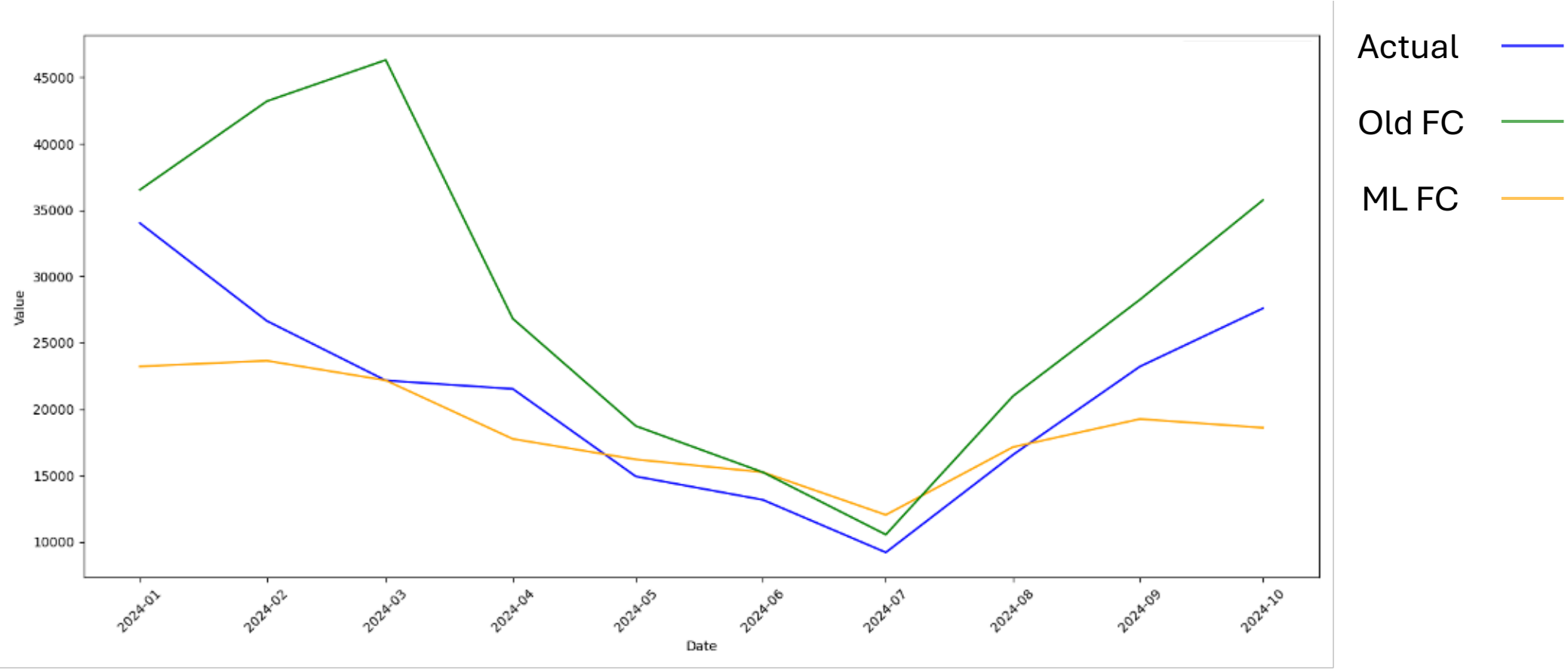


Elvenite

SCANDBIO

T+12 horizon

The ML model showed performance gains of up to **50 %** on average monthly



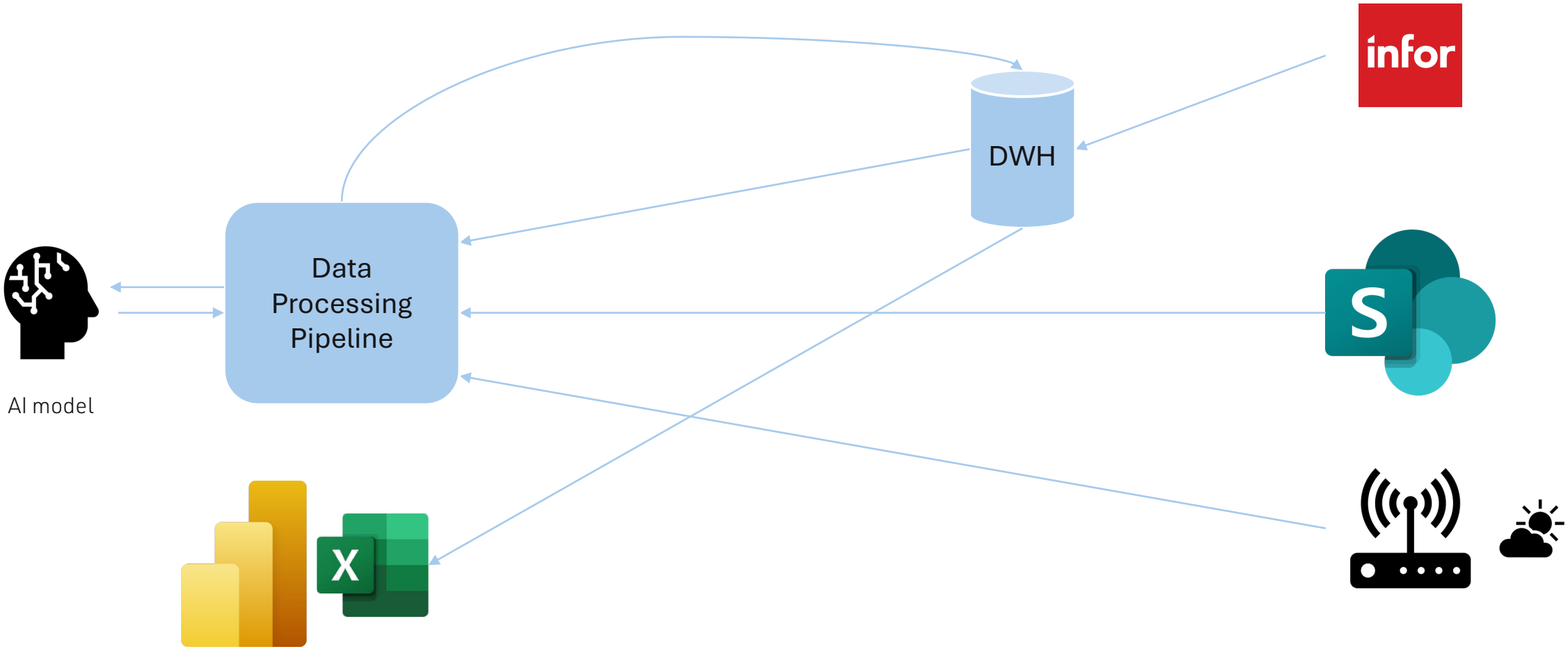
Elvenite

SCANDBIO

Ideas for valuable next steps

1. Continue to the next phase, implementing the forecast model to improve forecast accuracy and automate manual work
2. Future proof data processing from source systems and visualize data
    1. Build data pipelines from the source systems
    2. Create Business Intelligence reports of historical sales and forecasts to be able to monitor performance over time
3. Start looking into if the forecasts for the raw material can be improved and/or automated
4. Look into how the different forecasts can be combined, together with building digital twins of the sites to be able to optimize what, when, and where to produce products towards the customer base

How BI and AI can work together



Conclusions & reflections

Key success factors for a high-impact AI project

Involve the correct people: Make sure to involve the people working, closely with the data to be able to get to value creation quickly and not get stuck at question marks

Do not solely focus on the data, create an understanding of the business: In order to solve a company's challenges, you have to understand both their business and their data

Phase approach: Run projects as phases, first focusing on diagnosing your data before jumping straight into the AI treatment

Elvenite

SCANDBIO

Summary

- Scandbio phase a challenge where different customer segments and products exhibit different behaviour in terms of sales patterns
- Scandbio and Elvenite partnered up to perform a Data Proofing project together, looking into if Machine Learning can enhance their forecasts
- The project turned out successful and the outlook for Machine Learning powered forecasting solutions for Scandbio looks promising
- To run a successful AI project you need to involve the right people from the beginning, understanding both the business and the data, and make sure to diagnose your data before jumping straight into the AI treatment

Elvenite

SCANDBIO

SCANDBIO