Machine Learning applications at Teva Pharma with focus on forecasting

Laura Tolosi-Halacheva
Assoc. Director Data Science
laura.tolosi@gmail.com
Laura-Maria.Tolosi-Halacheva01@teva.bg

March 2024

The Pharma Industry

Teva in numbers

Enabling access to quality medicines and treatment options

The world's largest portfolio of medicines with more than

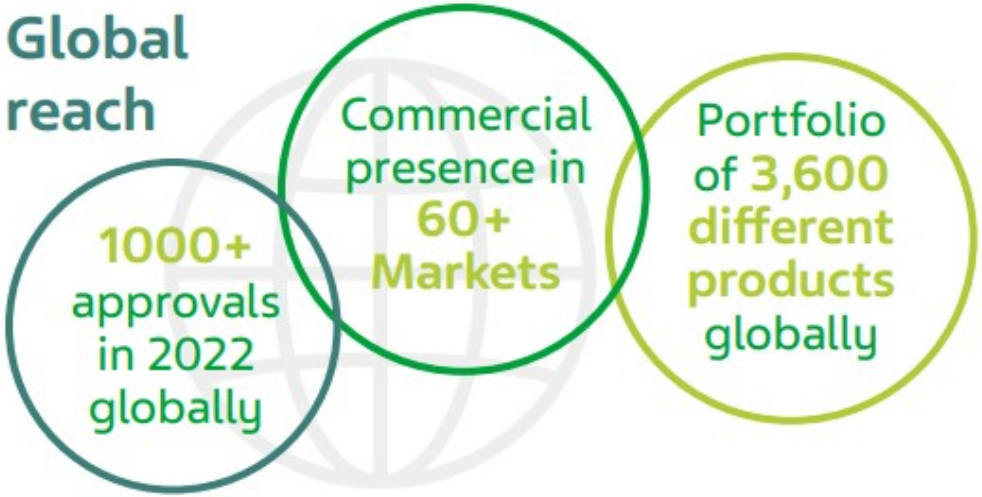
3,600 products

A full spectrum of products from generics, API and over the counter to innovative medicines and biologics

~ 37,000
employees
serving nearly
200,000,000
people every day

Teva in numbers

Global reach



Strong operational base and global infrastructure

53 production sites manufacturing

76 billion tablets & capsules annually

Scale - every second
10 Teva products are dispensed

Global generic leadership

Top 3 leadership position in over
25 markets

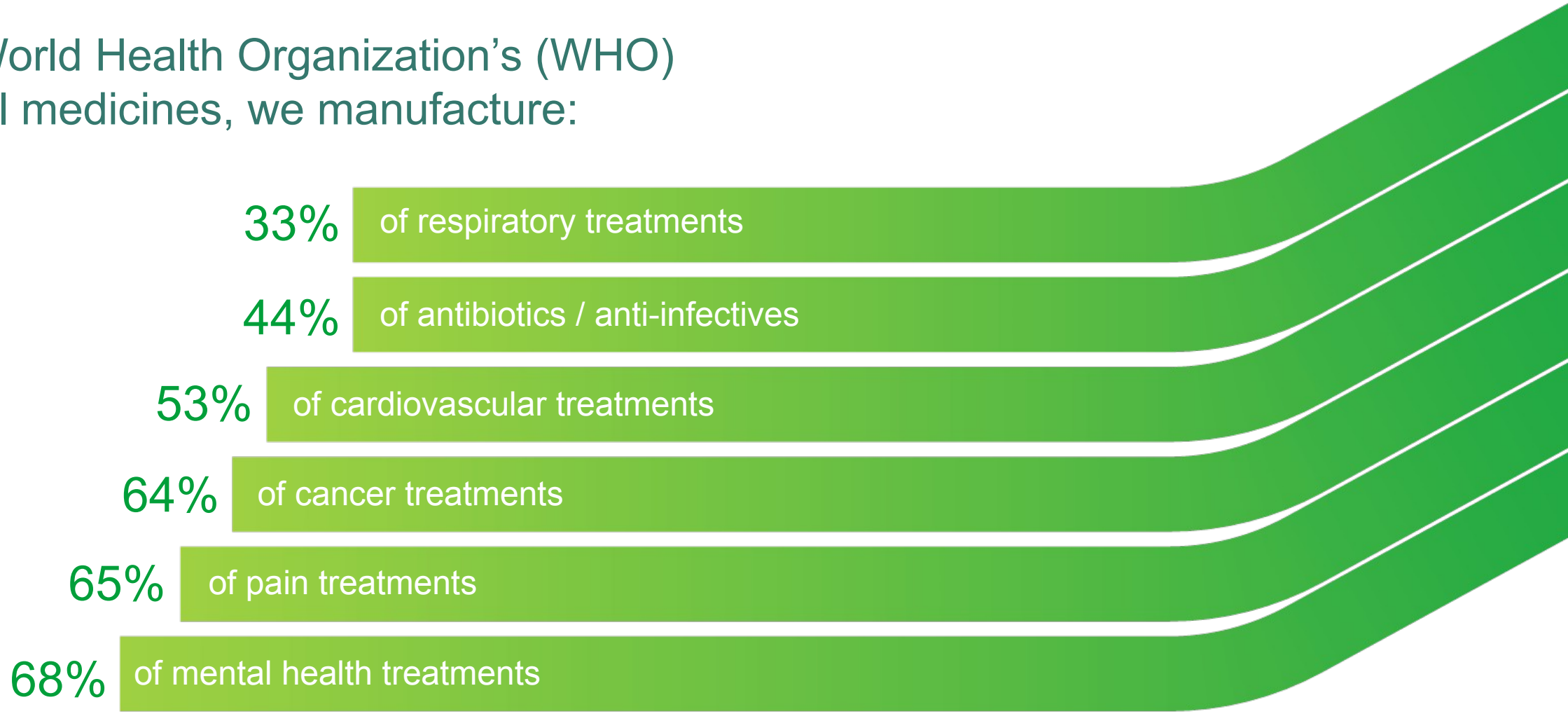
Delivered
$43.1B
in annual savings to healthcare systems in 14 of our markets in 2020

1 / 12 prescriptions in the U.S

1 / 5 prescriptions in the UK

Teva in numbers

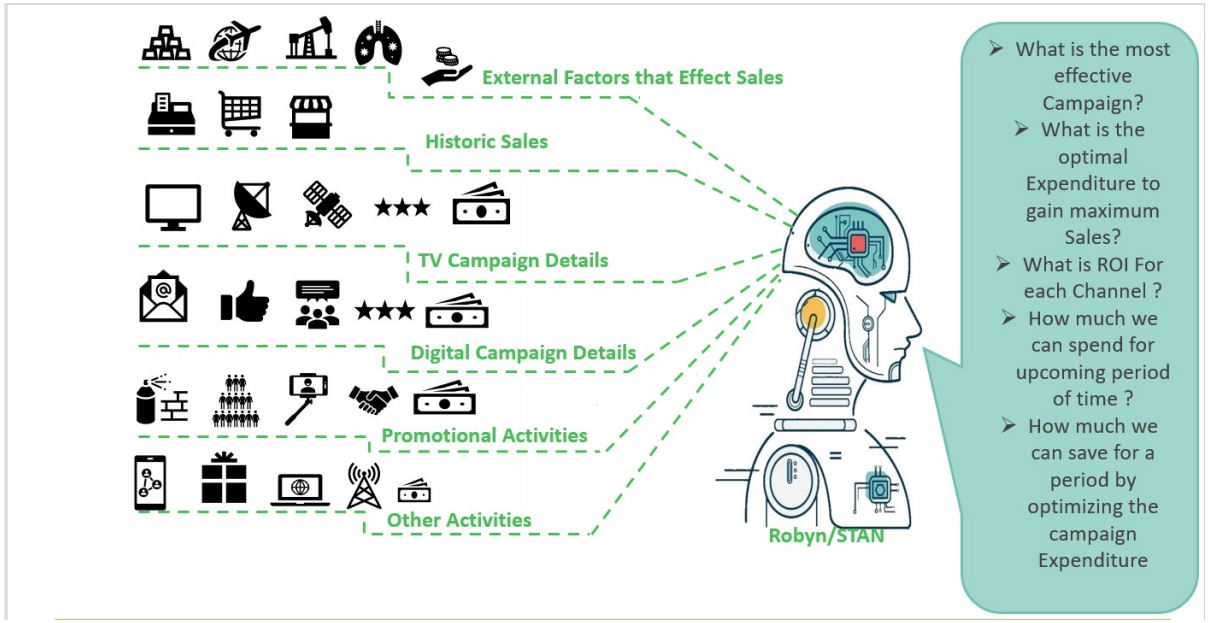
Of the World Health Organization's (WHO) essential medicines, we manufacture:

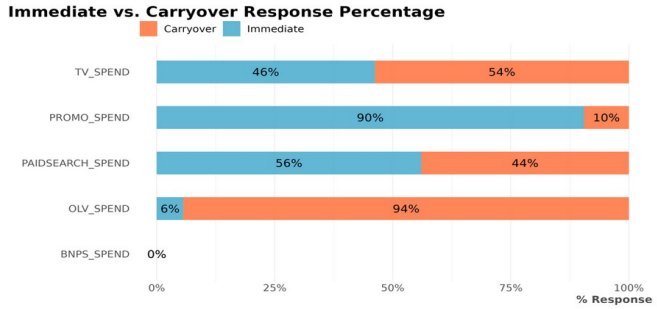


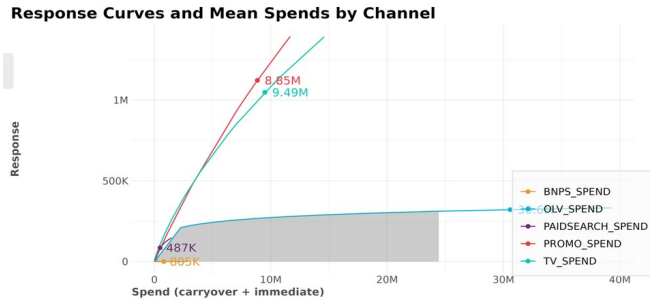
Source: IQVIA MIDAS MAT Q2 2019; Global = 32 countries
** Cancer excludes TCM; Diabetes excludes Insulin;*
***Anti-Infective includes antibiotics & antifungals;*
**** CVD = ATC1 C; Cancer = ATC1 L; Respi = ATC1 = R; Pain = ATC2 M1, N2; Mental Health = ATC2 N5, N6; Anti-Infective = ATC1 J; Diabetes = ATC2 A10 excluding ATC3 A10C, A10D, A10E*

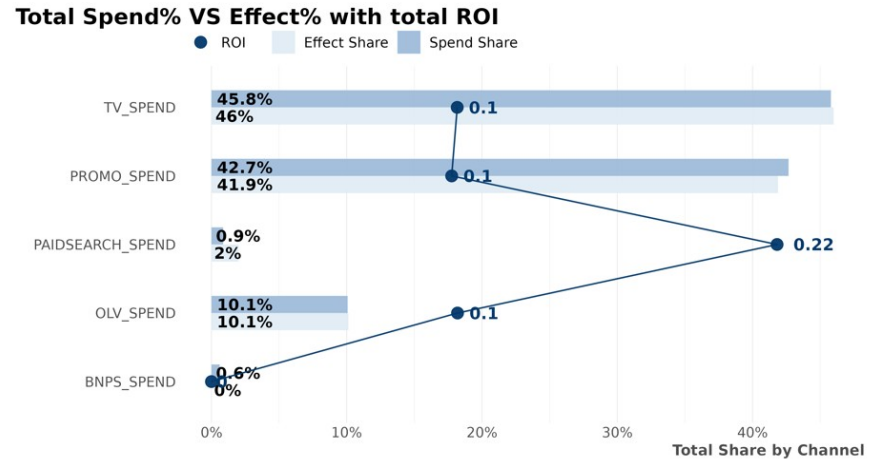

Data science
projects –
examples

Marketing Mix Modelling- saving an average of 10% on marketing budgets and 20% increasing the Efficiency of marketing Channel









- Use-case: Croatia TV commercial Campaigns for OTC products (eg. Olfen)
- Use-case: effective campaign for US Digihaler Market
- Use-case: Russian campaign optimization for Troxevasin, across multiple channels

Marketing Uplift

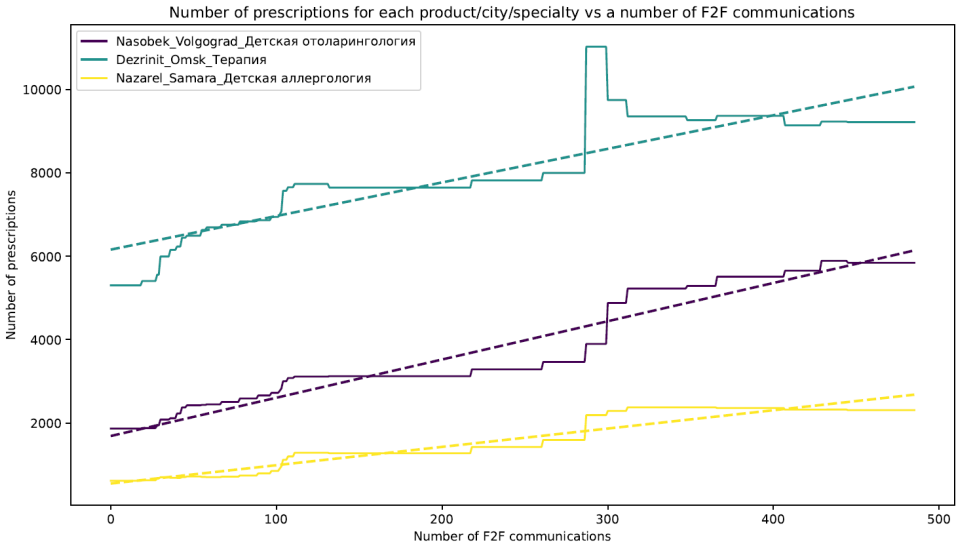
Goal:

Compute uplifts: a change in a number of prescriptions by increasing a number of communications by 1 for the channels with a linear dependence.

Results:

Computed uplifts for combinations of city and doctor’s specialty.

Example for 3 random samples in the F2F cha



Some technical information:

A master model is prepared using the CatBoost pa

Uplifts are least squares for predictions over a number of communications.

Canada Discount Optimization

Preferred Products

- **Target**: 6 Molecules, which already have good Market Share
- **Goal**: Increasing Profit by decreasing Discount Rates
- **Solution**: Recommend the smaller between median and individual Sales Rep's weighted average discount rate

```
 Chain  API NON-FORMULARY Molecule  OXYCOCET

You want to sell molecule  OXYCOCET  to chain:  API NON-FORMULARY  in  Nov/2023
         SalesRepName  mean   count
0  Ettahiri, Houceine    12  1803.0
1     Rahim, Mohammed    21  2022.0
2   Montuoro, Tiziana    21  1321.0
3    Ventura, Sherrie    34  3461.0
4      Dimech, Jessie    38  1266.0
5       Khan, Shameer    48  1242.0
6       Shisler, Mark    66  5788.0
7     Tessaro, Steven    69  3569.0
8     Joczys, Izabela    70  1724.0
Median: 38
Recommending the smaller between Median and Individual weighted mean discount
```

Focus Products

- **Target**: 39 Molecules with low Market Share
- **Goal**: Increasing Market Share by offering a bit higher Discount Rates
- **Solution**: ML model to maximize molecule Market Share with respect to Discount Rate

- Current solution for preferred products is deployed, sales reps are using it
- After 6 months of using the app, the discount for preferred dropped by 3%

Yield prediction

Pilot Product: Metformin

Project Overview

- Analyzing Yield Variability in Manufacturing, with a Focus on Tableting as a Critical Step Impacting Waste Generation

Objectives

- Identify the most important factors affecting yield
- Assess variations related to strength and seasonality of API manufacturing.
- Evaluate the manufacturer's impact on yield consistency.
- Explore the influence of raw material quality.
- Determine the impact of production variables on yield.

Data

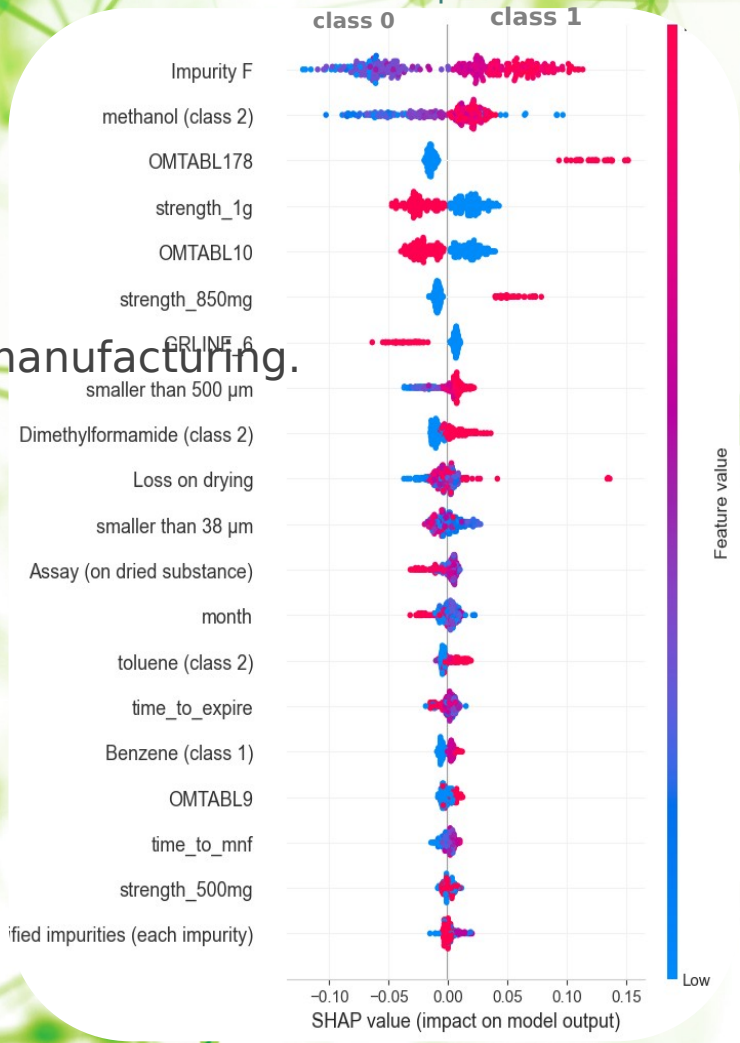
- Batches Produced in the period 2021-2023
- Raw Material and API Lab Results
- Activities Data for Granulation and Tableting Steps

EDA - Descriptive Statistics, Distributions and Correlations → Analysis of the variations of the data using PCA → Predictive Model (Random Forest Classification) for extracting Key Factors Impact → Next Steps: Adding IPC data to proceed with Root Cause Analysis

Findings

- Highest Yield Relations: Machine Type and Raw Material Quality
- Insufficient Data for Root Cause Analysis

Feature Importance

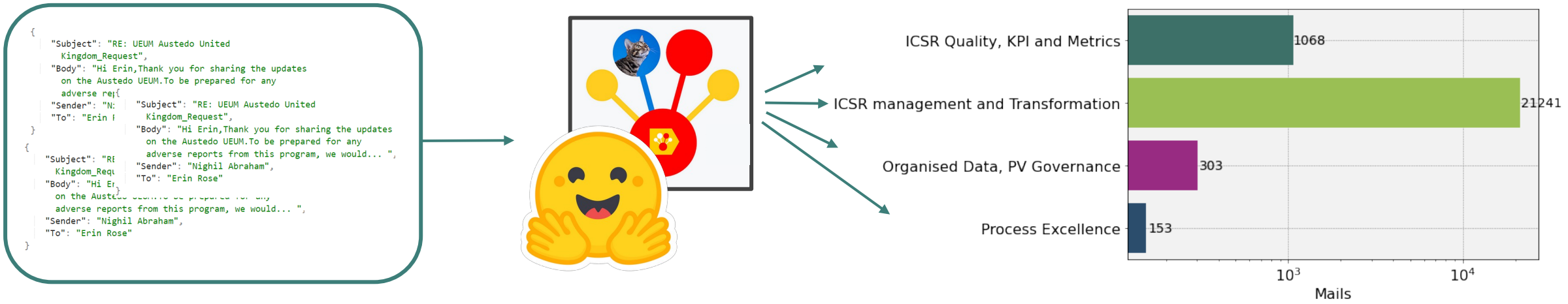


PV mail classification

PV is overwhelmed by emails in the shared mailbox "AEmailbox Teva"

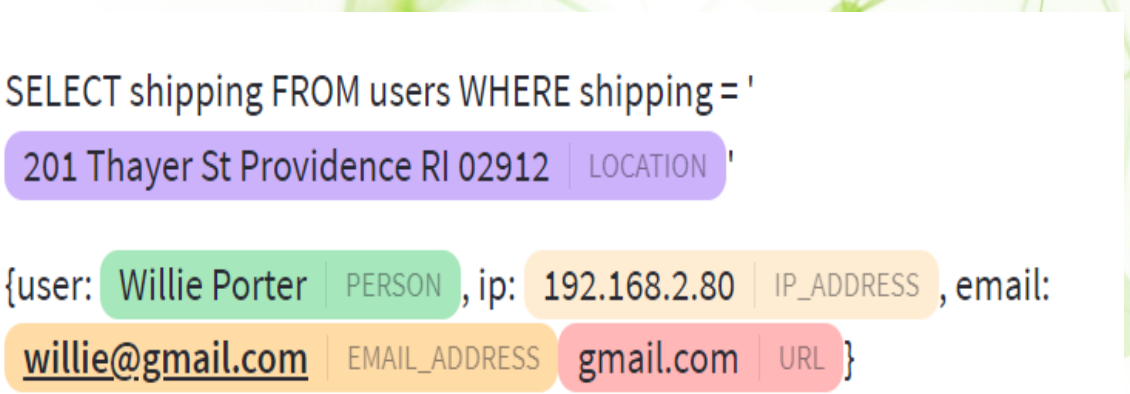
- ~10 000 emails monthly are currently classified by hand to 4 folders
- then there are 23 more subfolders to classify into

Our Goal: use state-of-the-art NLP and ML techniques to classify mails automatically



LSMI - PII

The project involves the development of a system to enhance privacy and data security in text documents based on guidelines. It focuses on identifying and removing personal information, such as names, phone numbers, addresses, and email addresses. These elements are then replaced with generic placeholders like [NAME], [PHONE], [ADDRESS], and [EMAIL]. This process ensures sensitive data is protected, maintaining confidentiality and adhering to privacy standards. The system's primary goal is to automate the safeguarding of personal information in texts, making it a valuable tool for data security management.



Anonymized

SELECT shipping FROM users WHERE shipping = '<LOCATION>'

{user: <PERSON>, ip: <IP_ADDRESS>, email: <EMAIL_ADDRESS>}

Canada Drug Out-of-Stock

Problem Statement

- Drug shortages are a prevalent issue in the pharma industry, often resulting in patient care disruptions and revenue losses for pharma companies.
- Accurately predicting drug shortages can enable companies to proactively manage their inventory and capitalize on market opportunities.

Proposed Solution

- We propose a time series forecasting approach to predict drug shortages.
- Here, we will analyze historical drug shortage data and additional data such as sales, to identify patterns that can be used to predict when our competitors are likely to go out of stock for a particular drug.

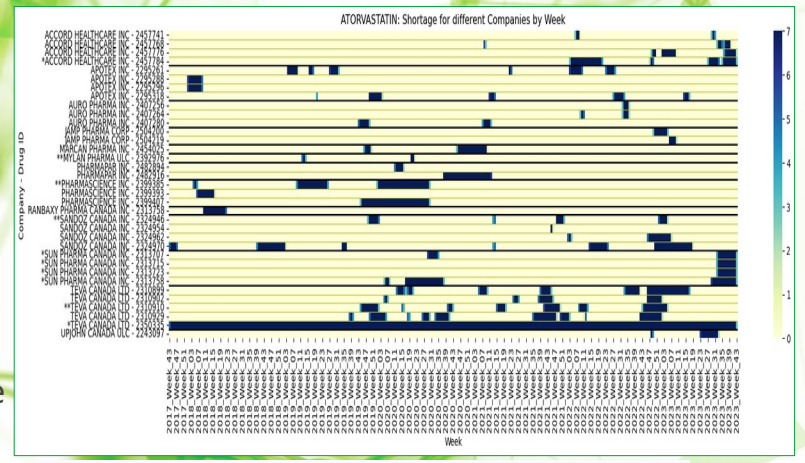


Fig. A sample drug shortage heatmap for competitors

Data Collection and Preprocessing

- To train and validate our time series models, we will utilize publicly available drug shortage data from drugshortagescanada.ca website. Additionally, we will incorporate proprietary IQVIA data.
- Before using the data, we will perform data preprocessing steps, such as data cleaning, normalization and feature engineering to ensure the data is of high quality and suitable for time series forecasting.

Model Development

- We will explore two different time series forecasting models: Gradient Boosting and Prophet.
- **Gradient Boosting** combines multiple weak models to create a powerful model.
- **Prophet** is specifically designed for handling seasonal and trend-based data.
- We will evaluate the performance of both models and select the one that provides most accurate predictions for our specific dataset.

Expected Outcome

- **Increased market share:** By capitalizing on our competitor stockouts, we can capture additional market share and grow our revenue.
- **Improved drug availability:** By proactively identifying potential drug shortages, we can maintain adequate inventory levels and ensure a consistent supply of drugs for our customers.
- **Enhanced customer satisfaction:** By maintaining drug availability and preventing stockouts, we can enhance customer satisfaction and loyalty.

API forecast

Goal:

Forecast sales of APIs for 12 months ahead.

Results:

A model for 82 products and a model for 31 products.

Quality with median metrics across the products:

Set of products	Model	NRMSE_mean, %	NRMDSE_median, %	NMAE_mean, %	NMDAE_median, %	MAPE, %	MDAPE, %	SMAPE, %	SMDAPE, %	INRSE, %	RAE, %
82 best predictable products	Model #1	74,43	56,15	65,42	55,77	200,91	52,02	69,06	56,88	119,22	113,76
	Dementra	90,28	61,43	67,71	60,90	118,85	58,86	80,60	63,08	166,11	157,42
	Baseline averaging	82,84	58,40	66,69	58,14	243,75	57,19	73,41	51,60	122,45	122,21
31 less predictable (extra) products	Model #2	70,90	52,87	62,56	52,21	152,87	46,16	65,15	49,52	112,35	117,20
	Dementra	82,36	54,50	54,88	54,46	69,42	53,73	60,79	46,40	156,26	140,30
	Baseline averaging	75,09	50,17	61,90	49,94	162,70	50,62	63,24	52,50	117,85	122,12
320 unpredictable products	Baseline averaging	269,25	Inf	238,45	Inf	Inf	Inf	153,34	181,39	184,09	108,00

* Baseline forecast: Average 24 months prior to the current month in the forecast (moving average).

** Monthly values are smoothed over 3 adjacent months (3 months rolling).

*** Not reliable metrics. Instead, SMAPE or SMDAPE are to be used.

All the metrics are scale-independent.

Sales forecasting

Questions

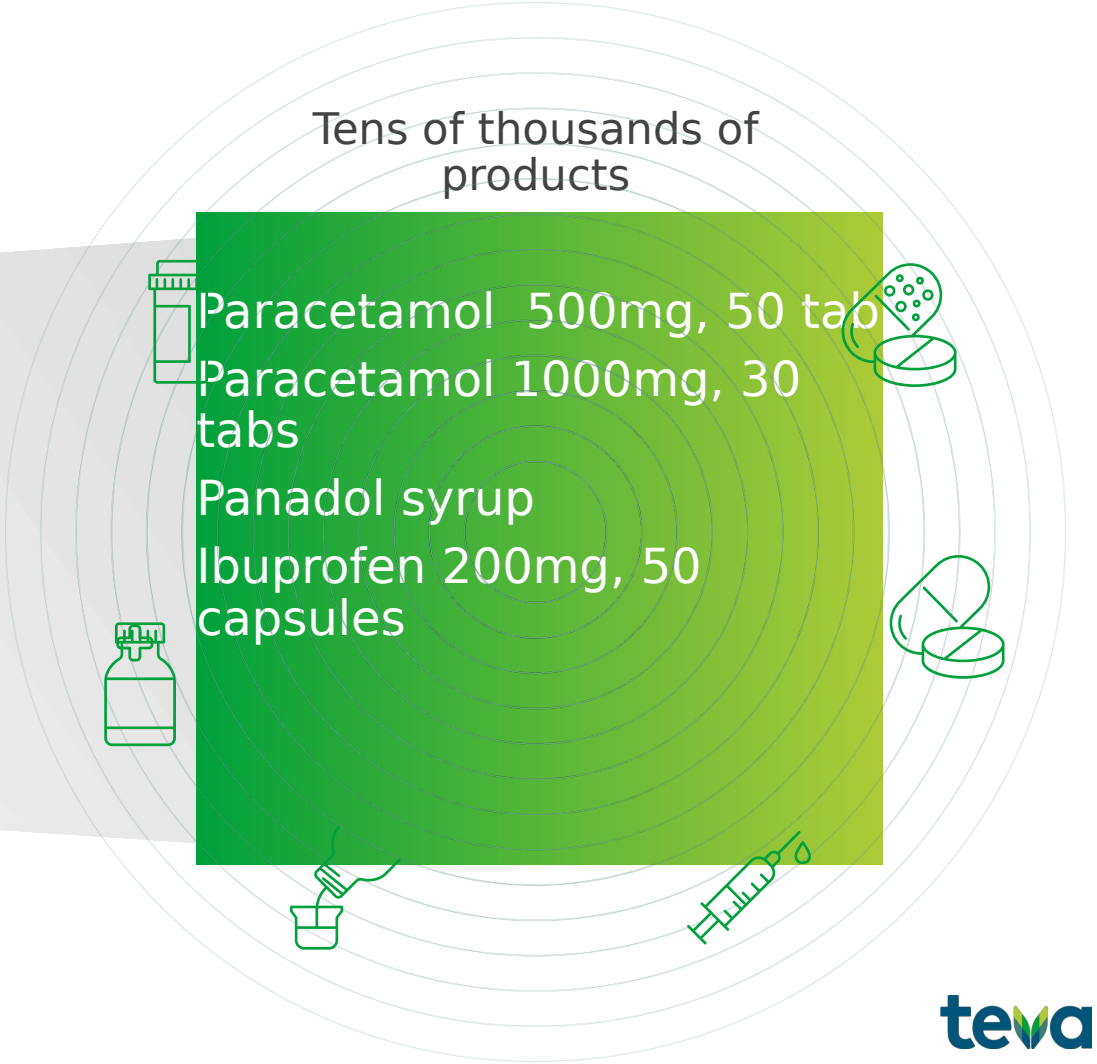
- What is TS forecasting?
- What is trend?
- What is seasonality?
- What are some models that you know for TS forecasting?
- What is stationarity?
- Packages for implementing TS forecasting in Python?
- How do we evaluate forecasts?
- Measures for evaluating TS forecasting?

Sales and stock depletion forecasting

Complexity: Teva’s portfolio is of tens of thousands of products

3600 molecules

Paracetamol
Ibuprofen
Valsartan

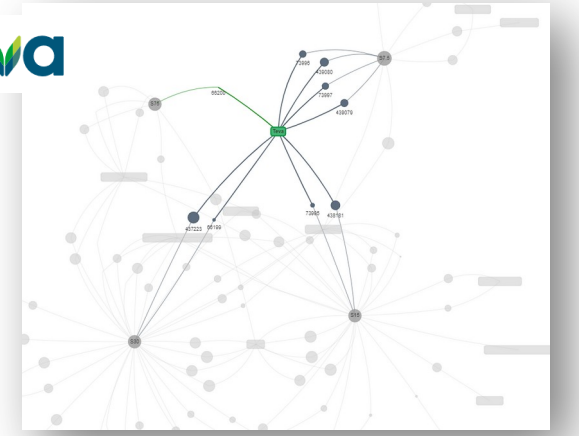


Forecasting medicine sales

- Maximize profit
- Inventory optimization:
  - avoid write-offs but do not miss opportunities
- High lead time, typically more than 6 mo
- Forecasts for 2 years ahead!
- Safety stock
  - We must keep a minimal amount of each product in stock
- Global manufacturing, supply chain
  - Each product can be manufactured in only 1-2 sites in the world
  - Packaging is specific for each market
- Demand is market-specific
- Demand is impacted by epidemics, competitors, marketing, etc

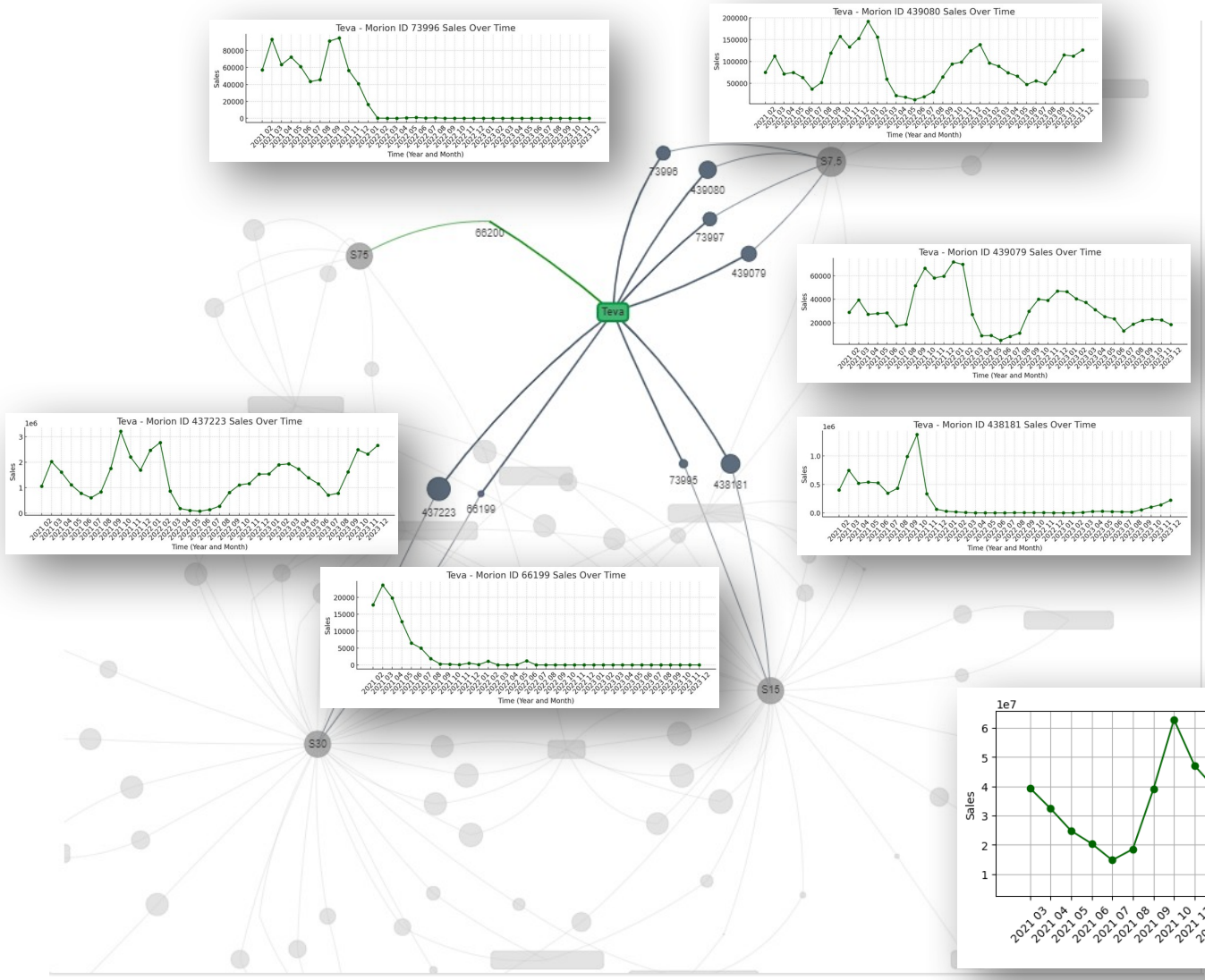
- Sales forecast must be at highest performance

Products network

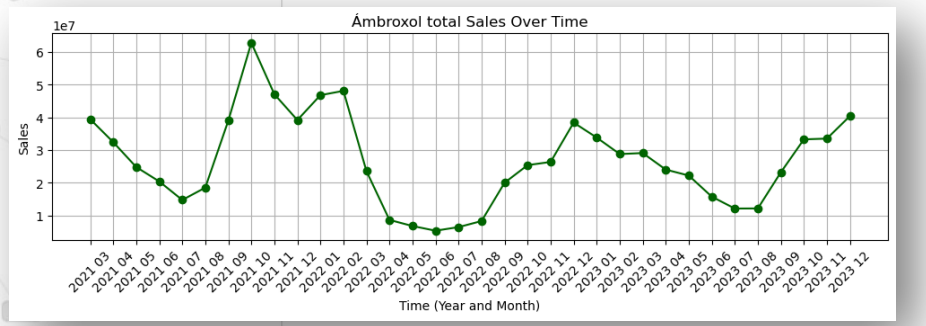


- Graph shows a competitive group, product is Ambroxol (cough syrup)
- There are many competitors producing various forms of ambroxol
- Teva sells 9 products, some more profitable than others
  - Products : grey bubbles
  - Profitability: size of the bubble
- Overall demand for Ambroxol in the market might be very predictable
- Preference for one product or the other is much more complex

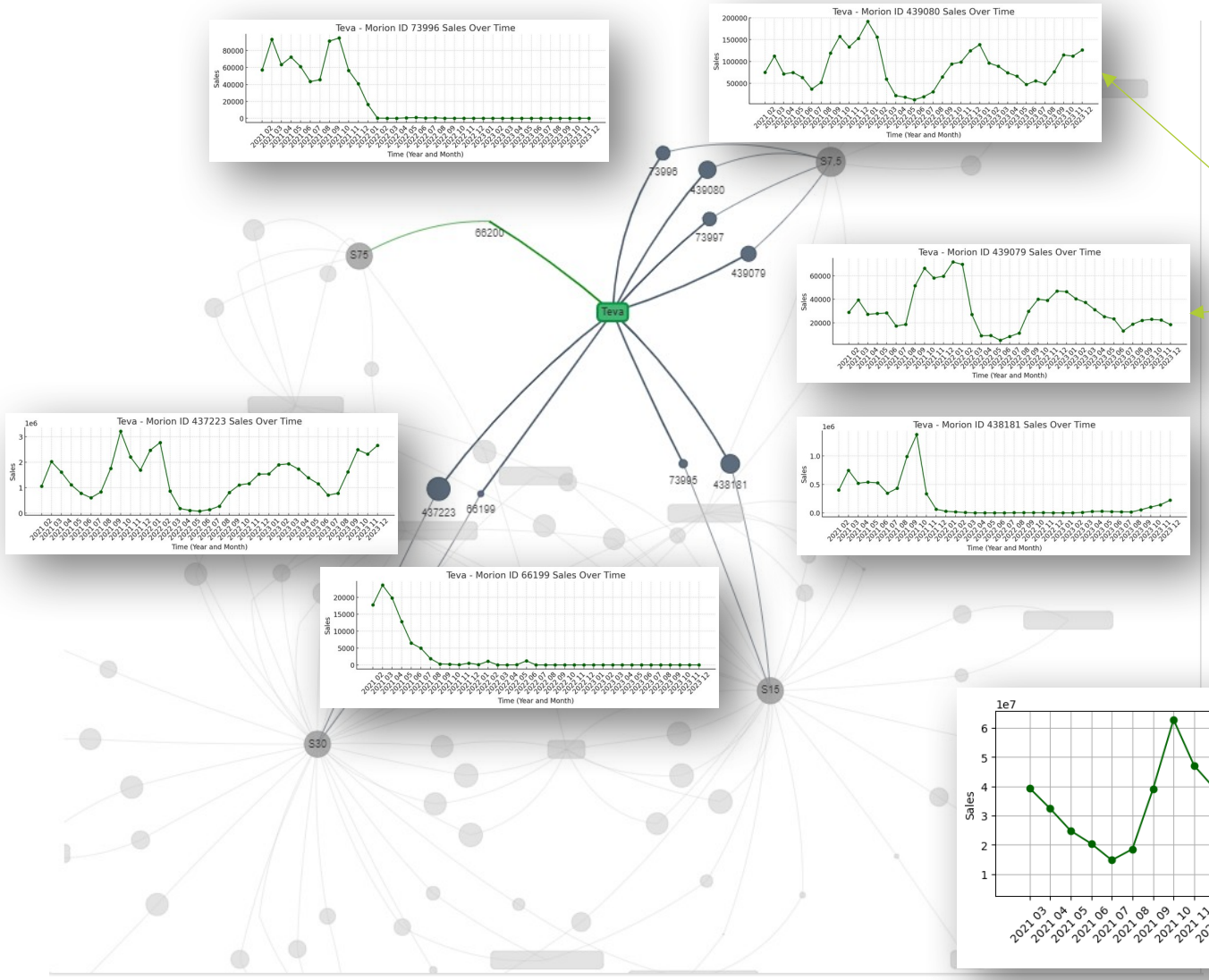
Teva products, overall sales



- There is seasonality:
  - Winter vs summer
- Covid peak in 2021
- Scale is different for different products
- There are discontinued or OOS products

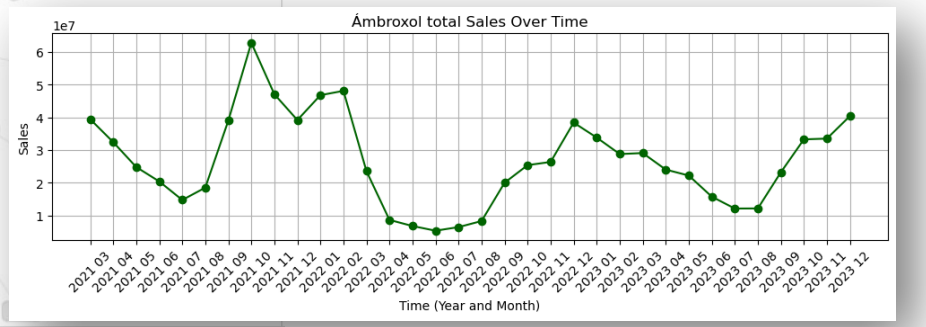


Teva products, overall sales

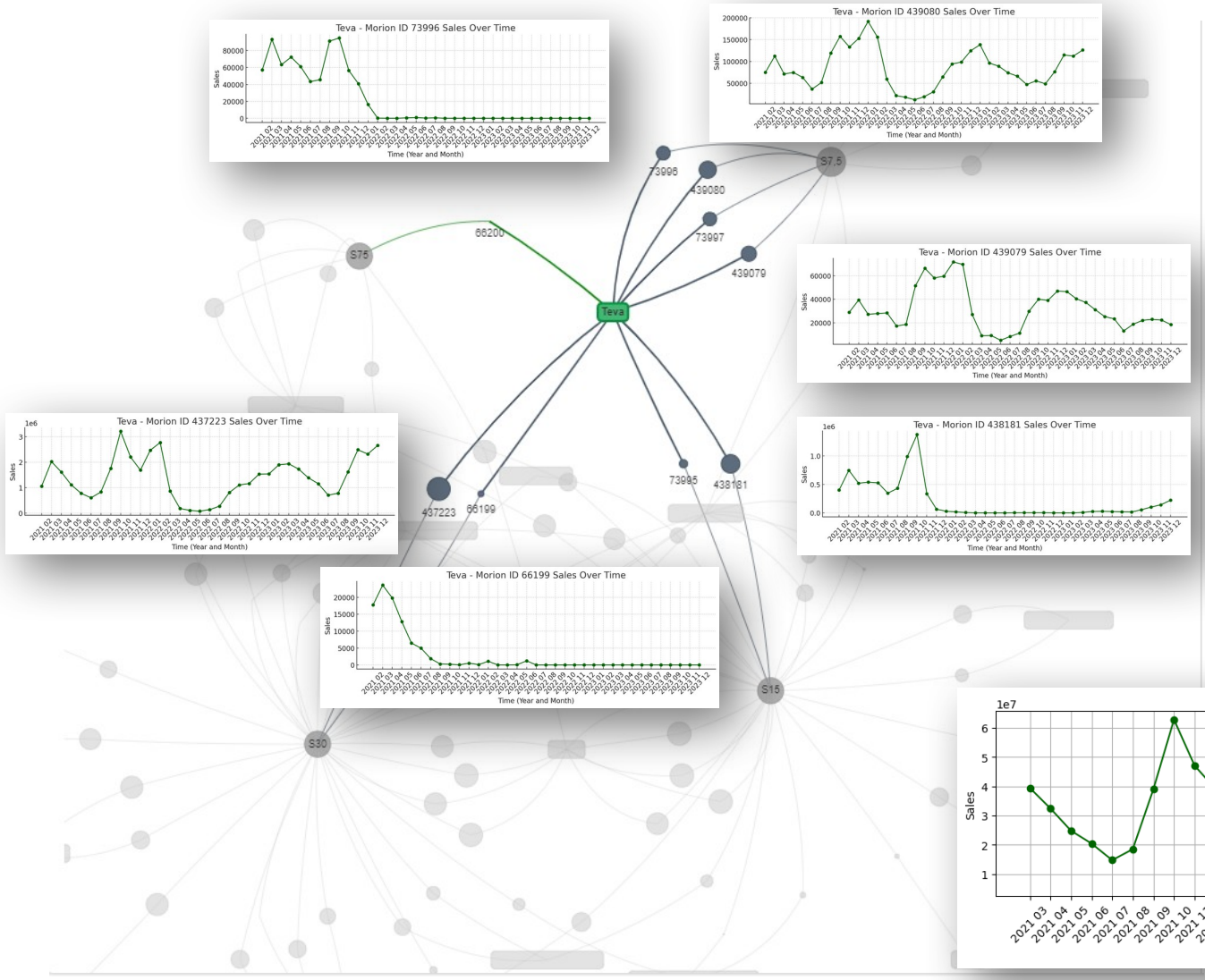


- Forecasting results with NBEATS model – Darts package

SMAPE	MAPE	Morion_ID
16	17	439079
61	92	439080
66	121	437223
99	233	438181
200	17667	73996
200	23034	73995
200	53660	66199
200	76662	73997



Teva products, overall sales

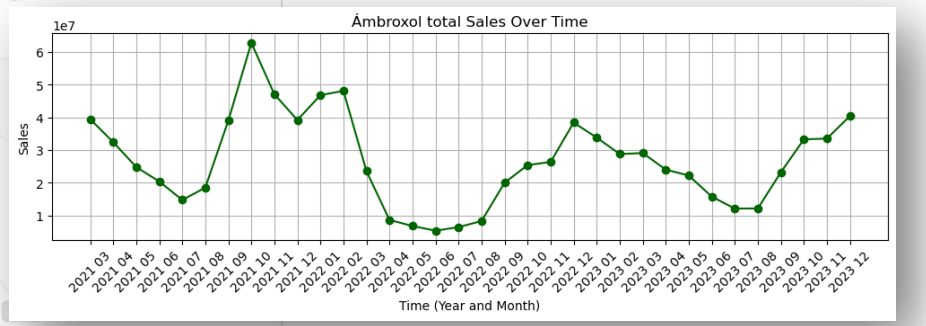


- MAPE:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

- SMAPE:

$$\text{SMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$



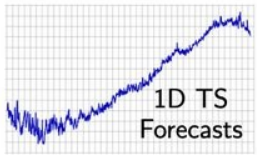
N-Beats: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting

- B. N. Oreshkin, D. Carpov, N. Chapados, Y. Bengio (2020) -

Can NNs Beat 1D Time Series Forecast Models?

Pure DL — Hybrid — Stats

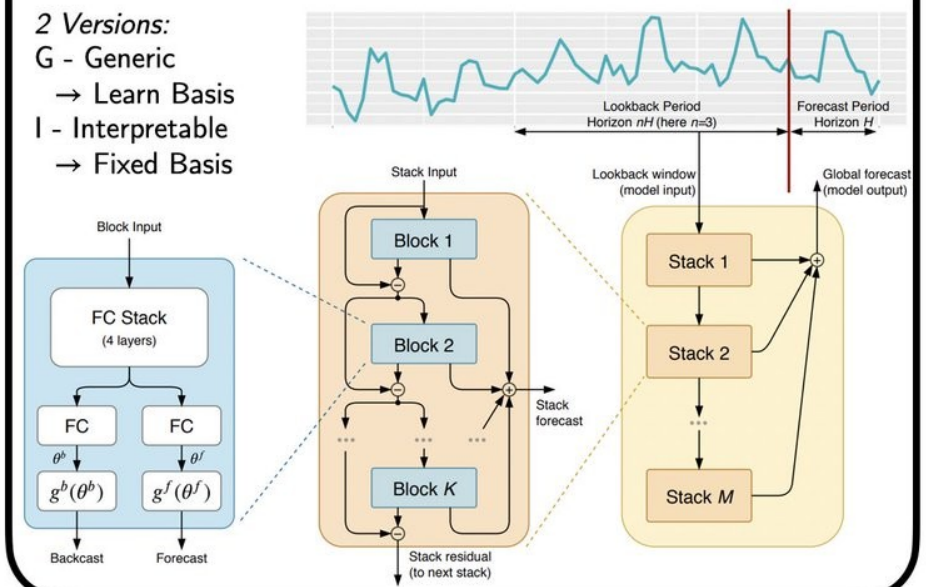
N-Beats — LSTM + Holt-Winters — ARIMA



⇒ N-Beats: ⇌ Residual Links & Deep Stack of FC Layers

N-Beats: Backcast & Forecast w. Deep Stacks

2 Versions:
G - Generic
→ Learn Basis
I - Interpretable
→ Fixed Basis

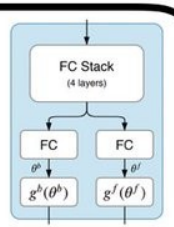


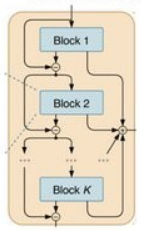
Basic Block [forecast/backcast]

$$\hat{y}_l = g_l^f(\theta_l^f) = \sum \theta_{l,i}^f \mathbf{v}_i^f$$

$$\hat{x}_l = g_l^b(\theta_l^b) = \sum \theta_{l,i}^b \mathbf{v}_i^b$$

→ $\mathbf{v}_i^f, \mathbf{v}_i^b$ learned/fixed basis





Residual Stacking

$$x_l = x_{l-1} - \hat{x}_{l-1}$$

→ Backcast = seq. input analysis

$$\hat{y} = \sum_l \hat{y}_l$$

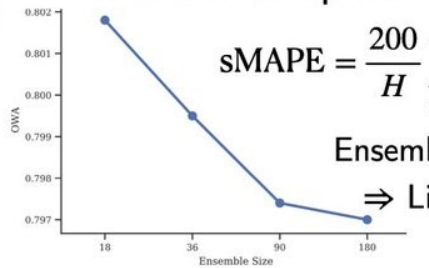
→ Final Forecast = sum of partial forecasts

Interpretable Constraints

Trend Stack: Polynomial $\hat{y}_{s,l} = \sum_{i=0}^{p} \theta_{s,l,i}^f t^i$

Season Stack: Fourier $\hat{y}_{s,l} = \sum_{i=0}^{\lfloor H/2-1 \rfloor} \theta_{s,l,i}^f \cos(2\pi i t) + \theta_{s,l,i+\lfloor H/2-1 \rfloor}^f \sin(2\pi i t)$

N-Beats Component Ablations



Ensemble: 10x Reduction
⇒ Little performance ↓

Stacks	sMAPE
1	11.154
3	11.061
9	10.998
18	10.950
30	10.937

Detrend	Seasonality	sMAPE
0	2	11.189
2	0	11.572
1	1	11.040
3	3	10.986

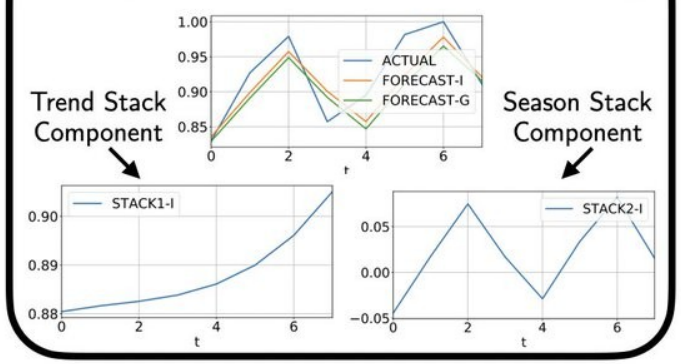
⇒ #stack saturate + works w. interpretability

N-Beats Ensembles > Stats Models & Hybrids

M4 Average (100,000)	sMAPE	OWA	M3 Average (3,003)	sMAPE	TOURISM Average (1,311)	MAPE
Pure ML	12.894	0.915	Comb S-H-D	13.52	ETS	20.88
Statistical	11.986	0.861	ForecastPro	13.19	Theta	20.88
ProLogistica	11.845	0.841	Theta	13.01	ForePro	19.84
ML/TS combination	11.720	0.838	DOTM	12.90	Stratometrics	19.52
DL/TS hybrid	11.374	0.821	EXP	12.71	LeeCBaker	19.35
N-BEATS-G	11.168	0.797		12.47		**18.47**
N-BEATS-I	11.174	0.798		12.43		18.97
N-BEATS-I+G	**11.135**	**0.795**		**12.37**		18.52

⇒ Ensemble: Different metrics, input windows, seeds

Interpretation: Trend & Seasonality



NN architecture for 1D TS forecasting ⇒ SOTA + Interpretable trend/seasonality components → Doubly residual stacking ≈ meta-learning?

* Figures adapted from Oreshkin et al. (2020)

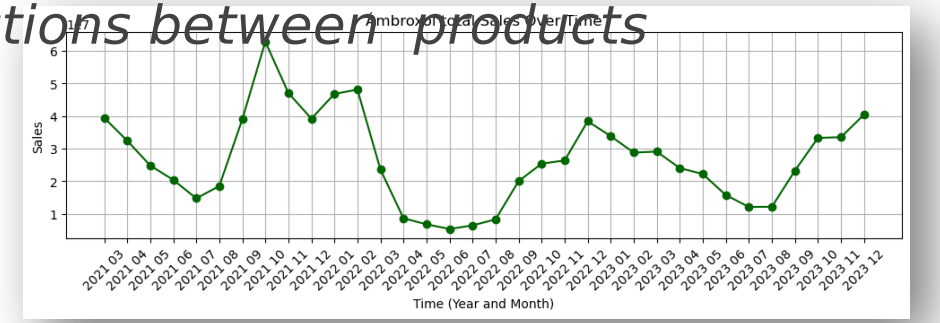
Teva products, overall sales

N-Beats : a very successful MLP architecture that achieved highest results at competitions

Q: what ts forecasting models do you know?

- N-Beats is operating with trend and seasonality, just like old ARIMAs, doing a better job on benchmarks and competition datasets
- It is interpretable
- Various improvements have been proposed already

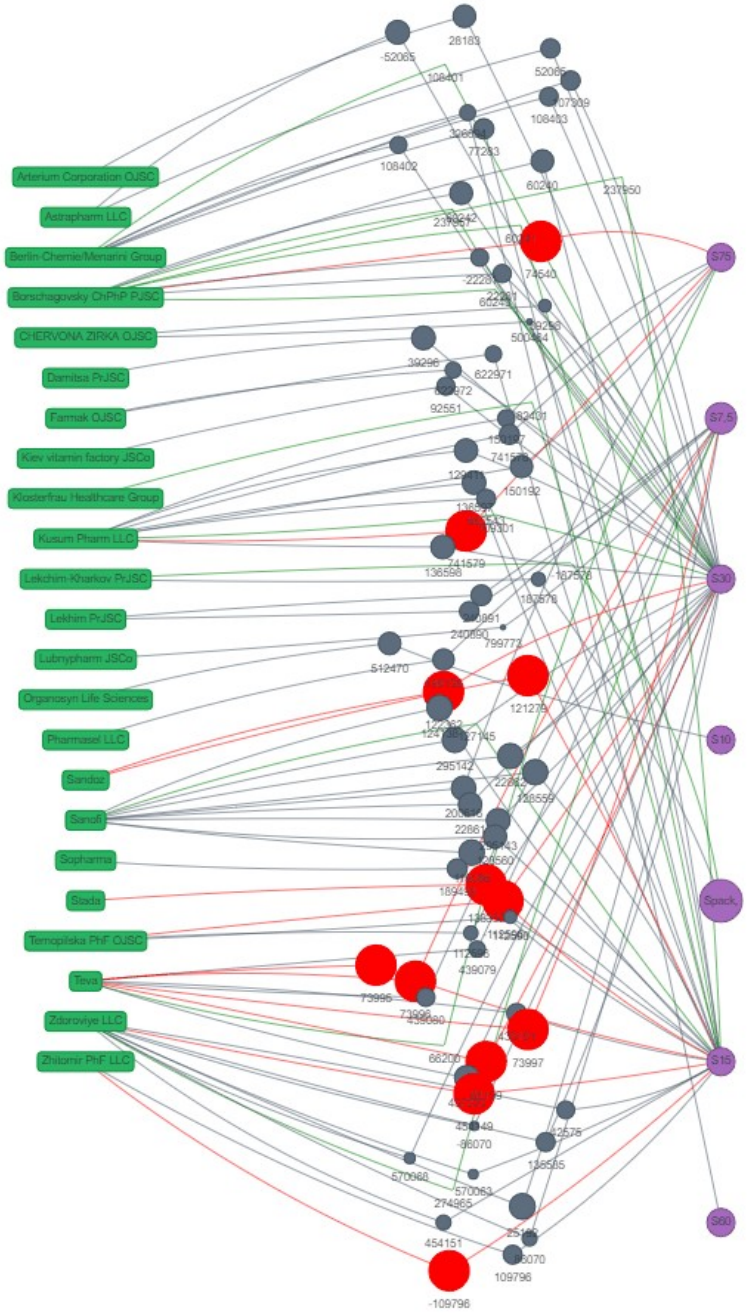
- It cannot forecast Covid and War in Ukraine:
- *It cannot consider the complex network interactions between products*




Stock depletion
event prediction

- Ambroxol goes out-of-stock often
- Red products went OOS
- Images shows one year of data, monthly

- Can we predict which product will go OOS next and for how long?
- Why do you think we want to predict that?



Let’s move to Canada

Drug Shortage:

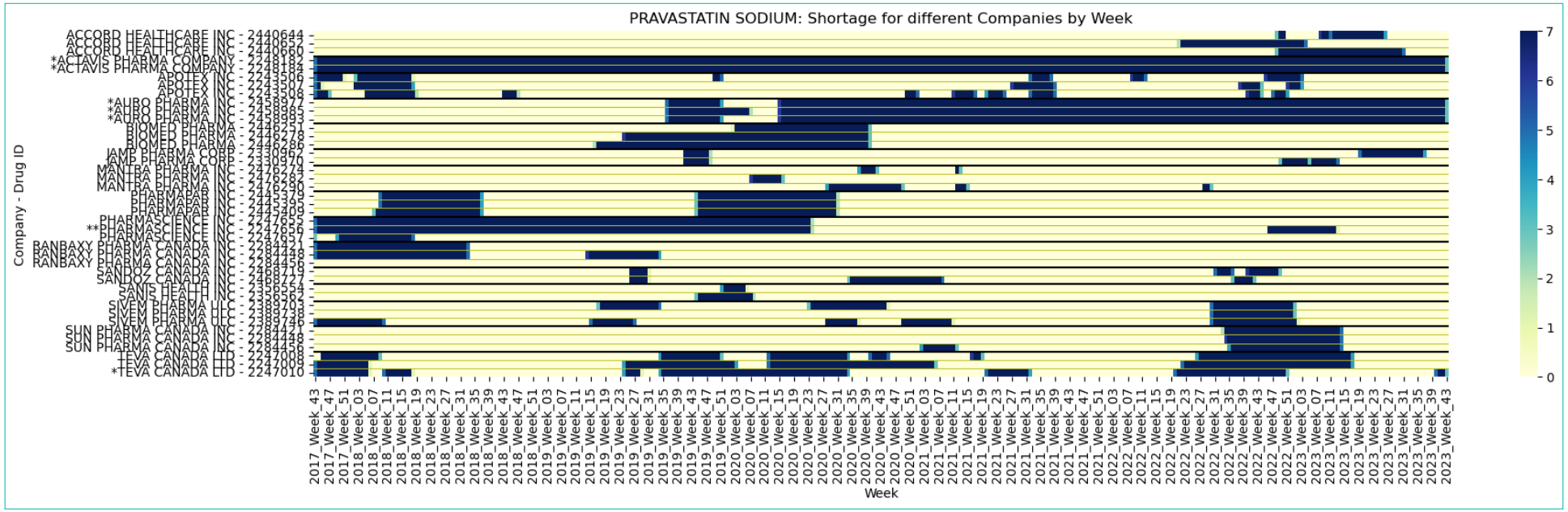
- Impact on patient care and healthcare costs.
- Delays in treatment, medication errors, patient deaths.

Drug Shortage Forecasting:

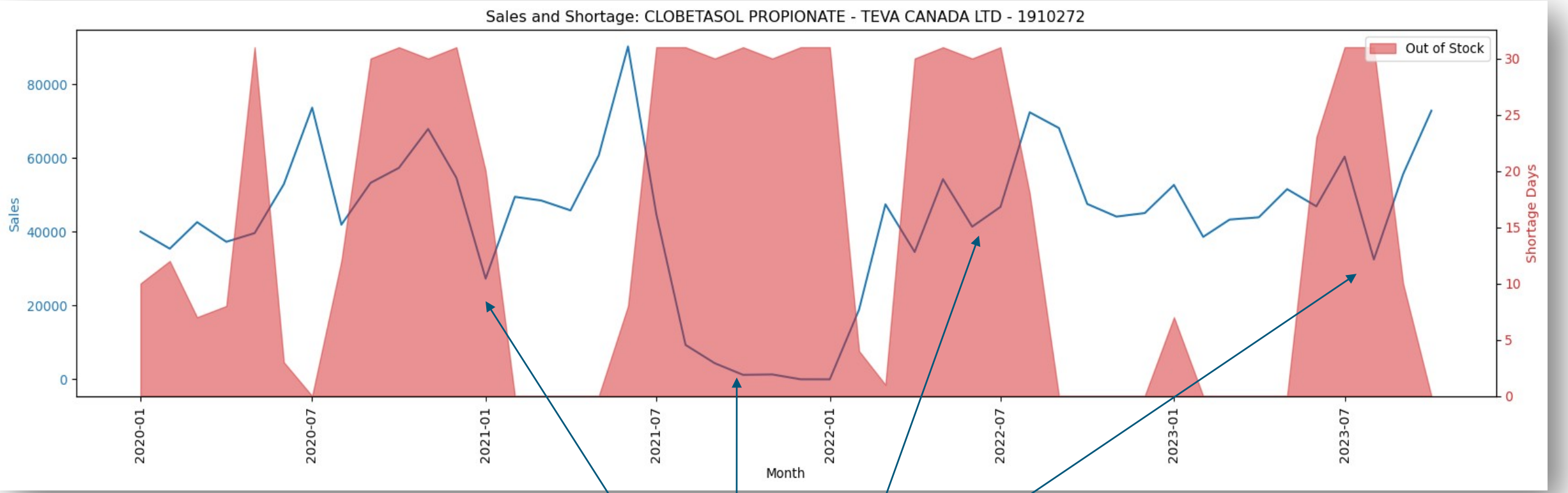
- Prevent and mitigate their impact,
- Opportunity to capture the market and sell more drugs.

Our Solution:

- Deep learning model to forecast drug shortages for a large dataset of approximately 4,100 DINs.
- The set includes only the ingredients that Teva manufactures.
- The goal of the model is to predict whether a drug will experience a shortage in the next 52 weeks, when exactly and for how long.

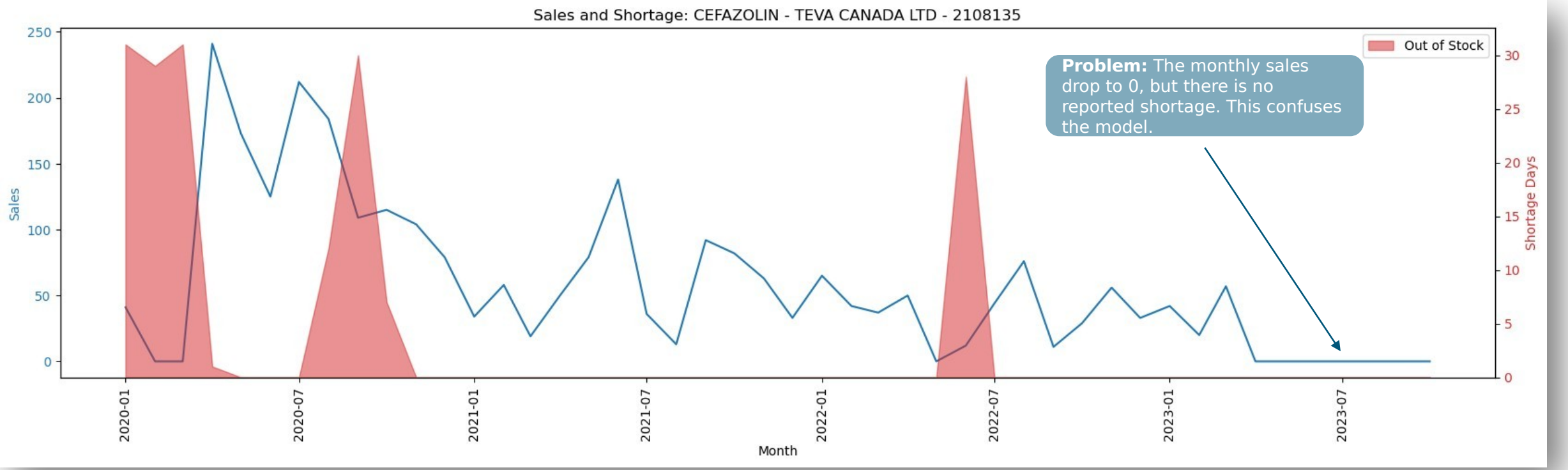


Visualization – Monthly Sales vs OOS



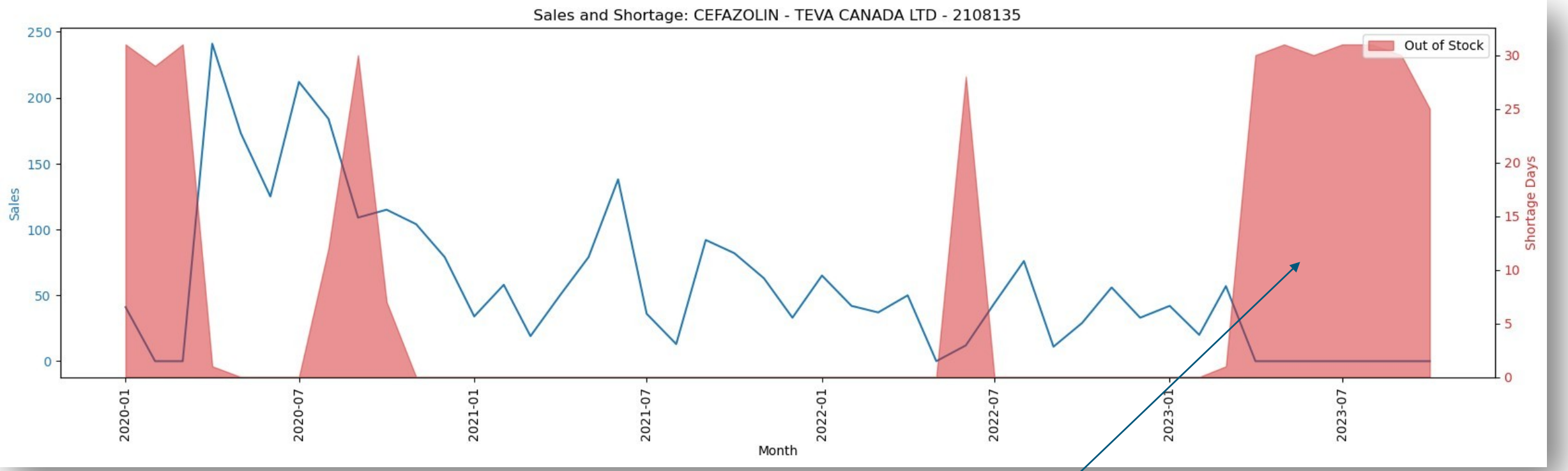
Ideal scenario: The monthly sales drop whenever there is a drug shortage.

Visualization – Monthly Sales vs OOS



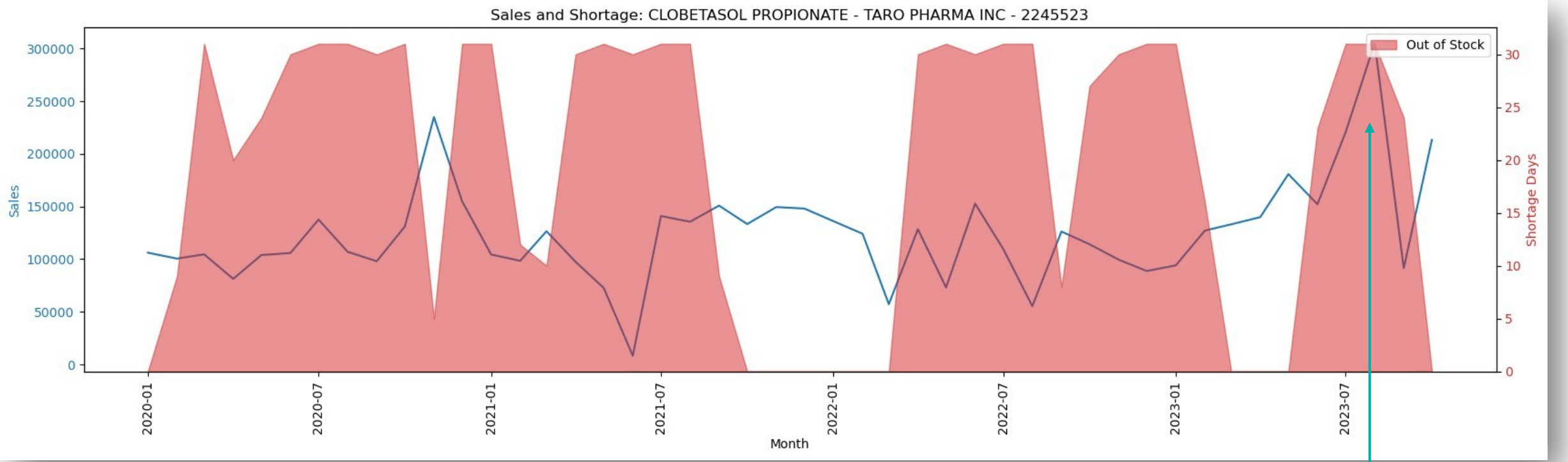
Solution: Use the discontinuation dataset to remove discontinued drug from dataset or mark them to be in shortage.

Visualization – Monthly Sales vs OOS



Plot after including the discontinuation dataset.

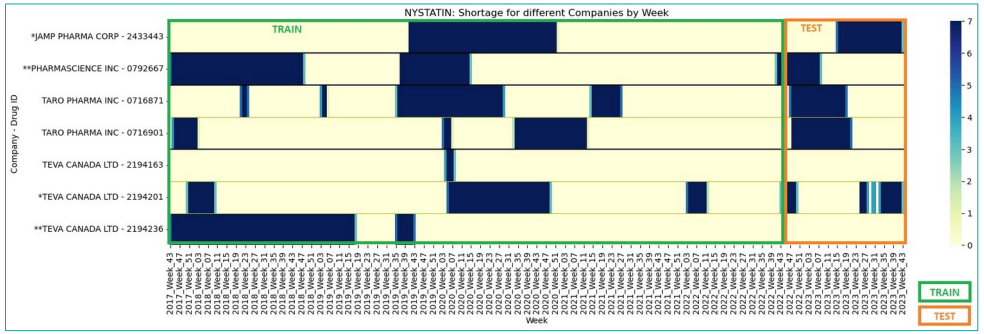
Visualization – Monthly Sales vs OOS



Noisy data: Sales go up even though shortage is reported.
Reason: One pack size is in shortage, but other pack size is still being sold.

ML solution

- Time series forecasting of binary values, 0 for no shortage, 1 for shortage, one value per week.



- The model relies on the past shortage patterns of all DINs simultaneously for forecast
- Trained: from Oct 2017 till Oct 2022
- Test: from Oct 2022 to Oct 2023 (1 year ahead) for all the DINs.
- 4153 DINs
- To evaluate the performance of our forecasting model, we employed a variety of metrics, including:
  - Accuracy: The overall proportion of forecasts (shortage and no shortage) that were correct.
  - Precision: The proportion of shortage forecasts that were actually correct.
  - Recall: The proportion of actual shortages that were correctly identified.

Model evaluation

How do we evaluate if the model is good?

	OOS	NOT OOS
Pred OOS	TP	,FP (false positives (cost is high
Pred NOT OOS	FN	TN

	Future week1	Future week2	Future week3	Future week4	Future week5	Future week6	FP	
Truth	0	1	1	1	0	0	–	
Forecast probability	0.3	0.4	0.9	0.7	0.6	0.55	–	← Model
Forecast OOS 1	0	0	1	1	1	1	2/6	← Thr = 0.5
Forecast OOS 2	0	0	1	1	0	0	0	← Thr = 0.7

Model evaluation

We are evaluating the performance of our model on an individual drug basis by calculating the metrics for each drug and then averaging across the entire dataset. This allows us to assess the model's generalizability across different drugs.

Sr No	Model	Forecast Period	Accuracy	Deviation ±	Precision	Deviation ±	Recall	Deviation ±
1	Recurrent Neural Network (RNN)	months 12	78.55	27.96	35.45	46.06	11.75	21.87
2	Long Short-Term Memory (LSTM) RNN	months 12	83.19	29.24	83.65	34.93	15.17	35.08
3	Gated Recurrent Unit (GRU) RNN	months 12	81.78	30.09	63.49	45.41	17.19	36.36
4	Block RNN	months 12	84.18	28.19	81.84	34.56	19.19	39.03
5	Block LSTM RNN	months 12	87.21	23.00	80.23	34.08	29.75	44.28
6	**Block GRU RNN**	months 12	84.74	24.23	**87.37**	28.10	20.36	34.19

Note: These results are calculated with a threshold of 0.7 (i.e. all predictions above 0.7 are considered as shortage)

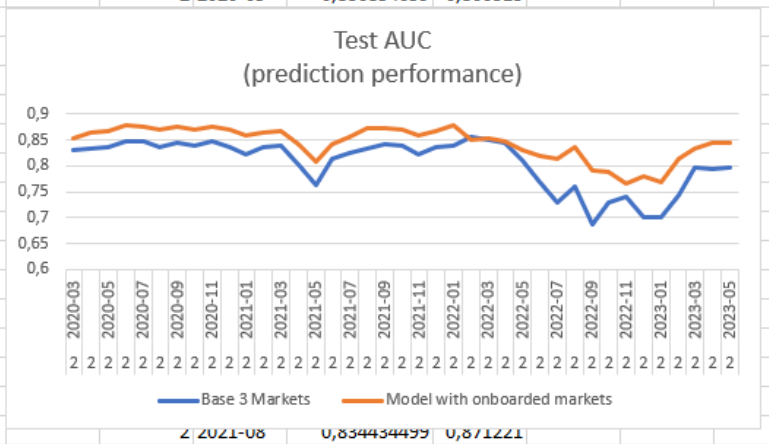
Let’s go back to Europe

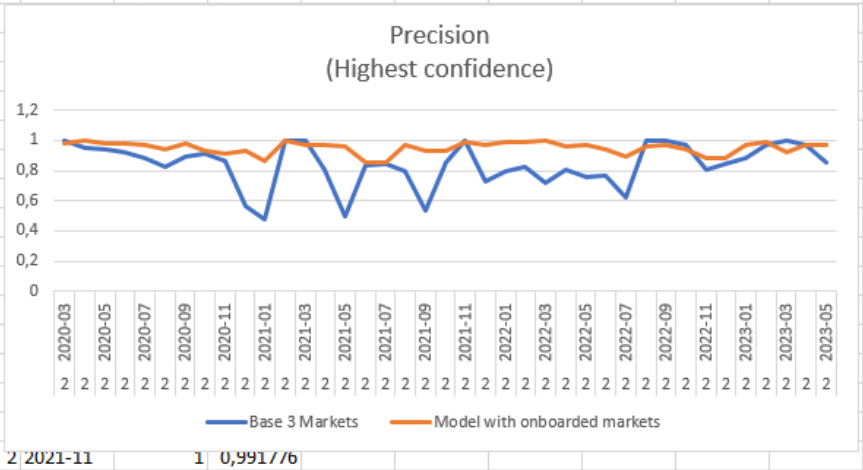
Cross-country OOS prediction:

- Predict Cross-Country OOS 6 months from present moment
- Markets: DE,DK,NL,FR,ES,HU,AT
- Products: 37 000 unique SKUs
- Vendors: 1024 unique vendors

- Vendors have cross-country strategies
- There is a network of competitors, within-market and cross-market
- We do not use it

	horizon	month	Base 3 Markets	Model with onboarded market	horizon	month	Base 3 Ma	
2	2	2020-03	0,829586983	0,853108	2	2020-03	1	0,976744
3	2	2020-04	0,833127588	0,863903	2	2020-04	0,948276	1
4	2	2020-05	0,836854659	0,866313	2	2020-05	0,944444	0,980341
5					2	2020-06	0,925532	0,982143
6					2	2020-07	0,885714	0,970179
19	2	2021-08	0,834434499	0,871221				
20	2	2021-09	0,842113969	0,872235				
21	2	2021-10	0,838516448	0,870069				
22	2	2021-11	0,822633377	0,857058	2	2021-11	1	0,991776





Conclusions

Advanced solutions - wish list

- Exploit the network-like underlying structure
  - Graph ML, time series on graphs, temporal graphs, signal processing on graphs
  - There are already transformers that can model ts on graphs
  - Attention is sparse, performance is not that good, black box
- Exploit multivariate aspects of time series data
  - Sales data, marketing data, stock depletion, disease incidence, economic factors, etc.

Challenges in Modeling Drug Shortage Events in the Pharmaceutical Domain

1st Laura-Maria Tolosi-Halacheva
Teva Pharmaceuticals
Sofia, Bulgaria
Laura-Maria.Tolosi-Halacheva01@teva.bg

2nd Radoslav Andreev
Teva Pharmaceuticalss
Sofia, Bulgaria
Radoslav.Andreev@teva.bg

3rd Oleg Shcherbakov
Teva Pharmaceuticals
Sofia, Bulgaria
Oleg.Shcherbakov@teva.bg

4th Eran Nevo
Teva Pharmaceuticals
Parsippany, USA
Eran.Nevo@teva.com

***Abstract*—Meeting demand for medicines translates into a very difficult manufacturing, supply chain and commercial optimization problem for every pharmaceutical company. Time series data reflecting missing stock, disease incidence, sales, marketing and others can help with better demand forecasting and supply planning. Accurate demand forecast ensures an optimal functioning of the healthcare environment, however modeling multivariate data in this domain remains challenging. In this article, we describe the typical data types and the modeling problems faced by the pharmaceutical domain regarding missing stock forecasting. We want to motivate the research community to turn their attention onto a set of problems that are highly complex, under explored, but at the same time critically relevant for a highly important domain.**

***Index Terms*—medicine demand, drug shortage, disease incidence, sales forecast, multivariate times series, transformers**

I. Introduction

The task of estimating and meeting demand for medicines faces challenges that are unique for the pharmaceutical domain. Difficulties arise from the large number of factors that impact medicine use in the population, such as disease incidence, healthcare practitioners' prescription trends, price, manufacturing cost, existing inventory and supply shortages, competition landscape, and more. Successful prediction strategies should utilize all these diverse data sources factors into account simultaneously, but most existing supply optimization algorithms are relying on sales data only [1] . Poor predictions have strong implications for the supply of medicines: over-forecasting causes oversupply of medicines which expire and need to be destroyed, under-predictions lead to lack of availability of medicines.

A straight-forward approach to modeling multivariate data is to engineer features from the multivariate time series and use boosting models for classification, regression, or downstream sequence prediction. Boosting has been shown to yield state-of-the-art performance on tabular data [24]. Such model is at least in theory superior to a univariate model, but the approach requires a large amount of manual effort for feature creation, and runs fast into computational issues due to a combinatorial explosion of features representing all pairwise interactions, cross-sectional and longitudinal.

There is hope that the recent AI advancements push the state of the art for this class of problems [5]. We feel that the research community can suggest creative approaches, including innovative neural network architectures that will discover dependency patterns implicitly without the need of feature engineering. In support of such initiatives, in this paper we describe typical data factors that play a role in medicine stock availability. We will formulate the most common modeling problems and the shortcomings of the current methods. The data sets that we use are proprietary, so we cannot disclose them in this work, but we try to describe their properties and interactions in enough detail.

Let's consider a universe of products $\{M_i\}_{i=1..N}$ which represent medicines. For simplicity, we will assume one medicine is one *molecule*. For example, the widely known ibuprofen, paracetamol, valsartan, acyclovir are molecules. They have certain proven therapeutic properties and they are used to treat a specific set of conditions.

For each molecule in the universe, we can gather a wide range of related dynamic factors that can be modeled together. This way, we have a *multivariate time series* level under the multiple time-series modeling problem. We firmly believe that leveraging multivariate characteristics can lead to high-accuracy modeling of stock depletion.

The paper is structured as follows: Section II describes Out of Stock (OOS) events data. Section III presents a list of factors that can be combined with the missing stock data for more accurate modeling. Section IV describes several relevant modeling tasks and section V concludes the paper.

II. Missing Stock Events in the Pharma Industry

Medical products are frequently out of stock, affecting the lives of patients that depend on them. There are many reasons for stock depletion, including lack of active ingredients, manufacturing disruptions, logistic difficulties, compliance issues and increased demand. Missing stock cannot be replenished

Paper accepted for Multisa 2024