



3rd edition

The **financial challenge** of the year

NOVARTIS DATATHON

online



3rd edition

The financial challenge of the year
NOVARTIS DATATHON
online

TEAM 25

(the Team Formerly Known as Adevinta Transformers)

Team member Xavier Belda

Team member Didac Fortuny



Adevinta
Spain

fotocasa habitaclia Infojobs
coches.net motos.net milanuncios





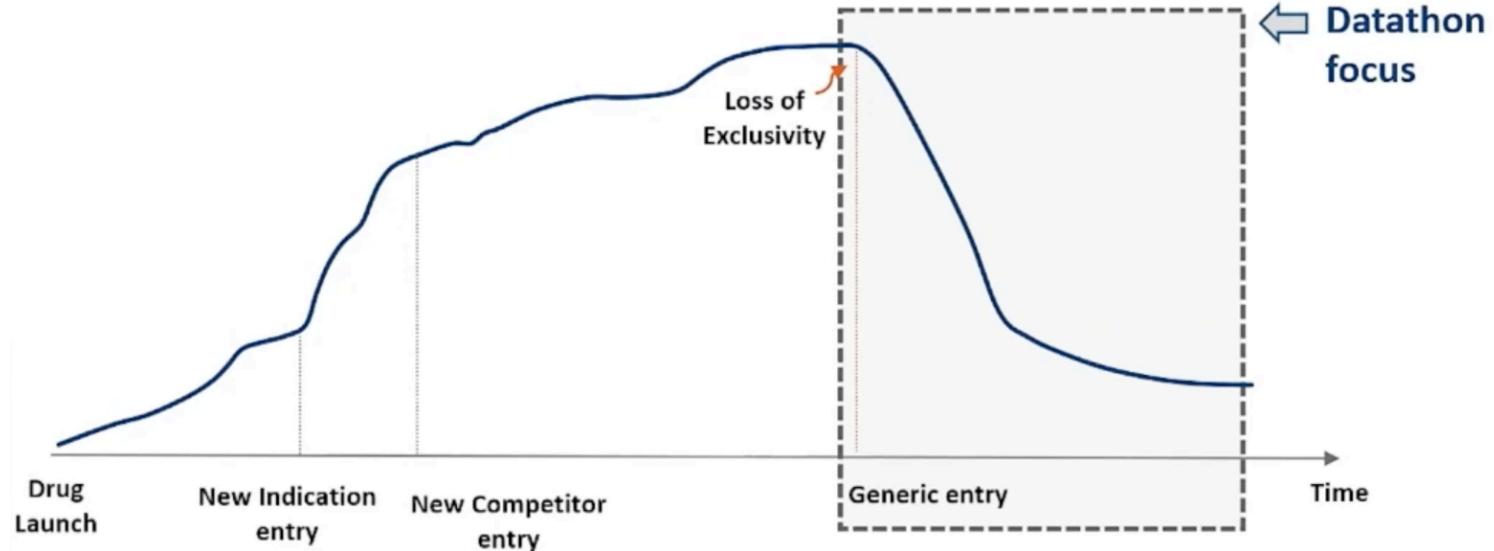
3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Background





3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Challenge

1. Data Science challenge

Participants are required to provide **24 months of volume forecast** after the generic entry date together with a **95% confidence intervals** for the given prediction for all the brands in the test set.

2. Business challenge

All teams that present in front of the Jury will be asked to provide a **deep exploratory analysis** on the correlation **between features** provided and the **impact in the volume sold** after the gx entry. We encourage the participants to use visualization tools.



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Data

- **Target:**

- **Volume ***: Historical (pre-gx) volume for **1078 country-brands** that went generic in the past.
- **Train: 887 observations** for which in addition are provided 24 months of volume after the gx entry date.
- **Test: 191 observations** for which a forecast needs to be provided for month_num = 0 (month of gx entry) to month_num = 23 after the gx entry date.

- **Features:**

- **Therapeutic Area**
- **Package**
- **Panel (Channel of Distribution)**
- **Number of gx**

*Volume can be in different units depending on the country and brand (miligrams, packs, pills, etc.)



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Metrics

$$\begin{aligned} PE_j = & \quad 0.5 \cdot \left(\frac{\sum_{i=0}^{23} |Y_{j,i}^{act} - Y_{j,i}^{pred}|}{24 \cdot Avg_j} \right) + 0.3 \cdot \left(\frac{|\sum_{i=0}^5 Y_{j,i}^{act} - \sum_{i=0}^5 Y_{j,i}^{pred}|}{6 \cdot Avg_j} \right) \\ & + 0.1 \cdot \left(\frac{|\sum_{i=6}^{11} Y_{j,i}^{act} - \sum_{i=6}^{11} Y_{j,i}^{pred}|}{6 \cdot Avg_j} \right) + 0.1 \cdot \left(\frac{|\sum_{i=12}^{23} Y_{j,i}^{act} - \sum_{i=12}^{23} Y_{j,i}^{pred}|}{12 \cdot Avg_j} \right) \end{aligned}$$

Finally the Prediction Error PE will be the average across all the prediction errors PE_j of all brands n in the test set:

$$PE = \frac{1}{n} \sum_{i=1}^n PE_j$$



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

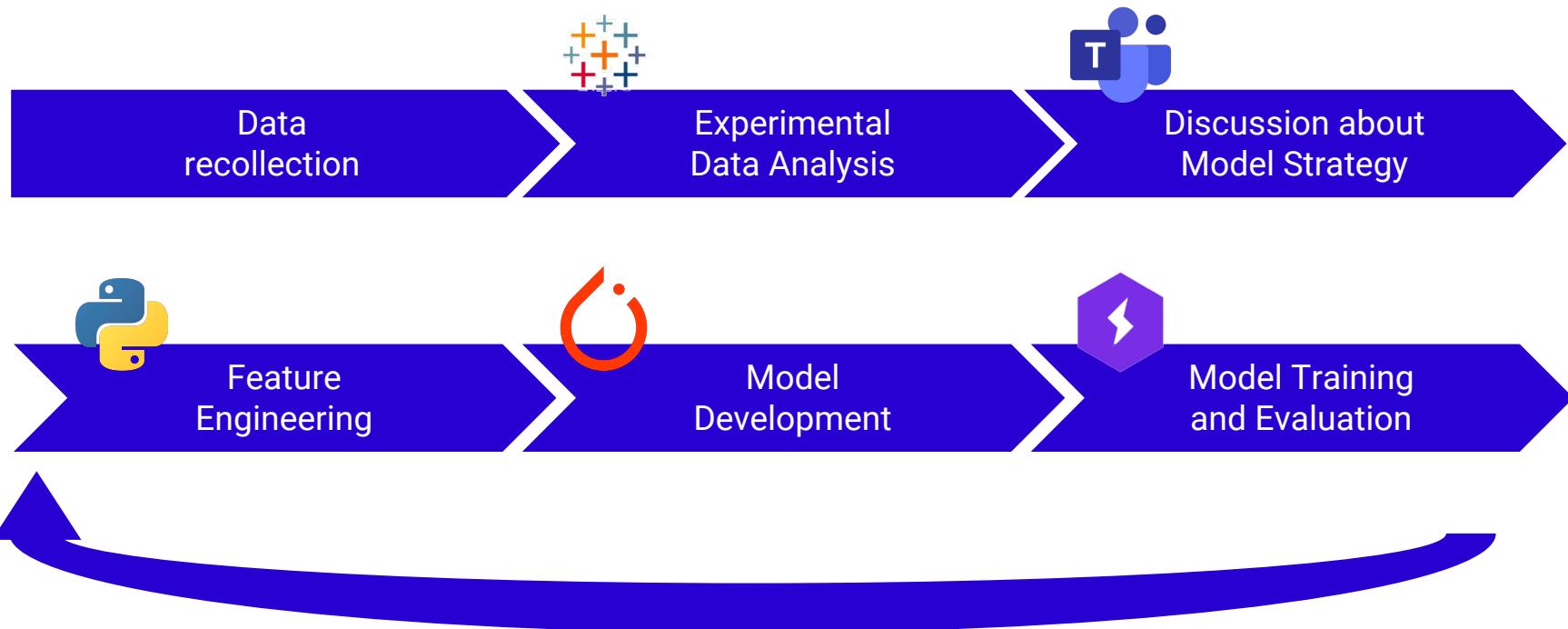
Metrics

$$\begin{aligned} CE_j = & \frac{0.6 \cdot \left[\sum_{i=0}^5 (0.85 \cdot |U_{j,i} - L_{j,i}| + 0.15 \cdot [\frac{2}{0.05} \cdot (L_{j,i} - Y_{j,i}^{act}) \mathbf{1}\{Y_{j,i}^{act} < L_{j,i}\} + \frac{2}{0.05} \cdot (Y_{j,i}^{act} - U_{j,i}) \mathbf{1}\{Y_{j,i}^{act} > U_{j,i}\}]) \right]}{6 \cdot Avg_j} \\ & + 0.4 \cdot \left[\frac{\sum_{i=6}^{23} (0.85 \cdot |U_{j,i} - L_{j,i}| + 0.15 \cdot [\frac{2}{0.05} \cdot (L_{j,i} - Y_{j,i}^{act}) \mathbf{1}\{Y_{j,i}^{act} < L_{j,i}\} + \frac{2}{0.05} \cdot (Y_{j,i}^{act} - U_{j,i}) \mathbf{1}\{Y_{j,i}^{act} > U_{j,i}\}])}{18 \cdot Avg_j} \right] \end{aligned}$$

Finally the Confidence Error CE will be the average across all the confidence errors CE_j of all brands n in the test set:

$$CE = \frac{1}{n} \sum_{i=1}^n CE_j$$

Roadmap





3rd edition

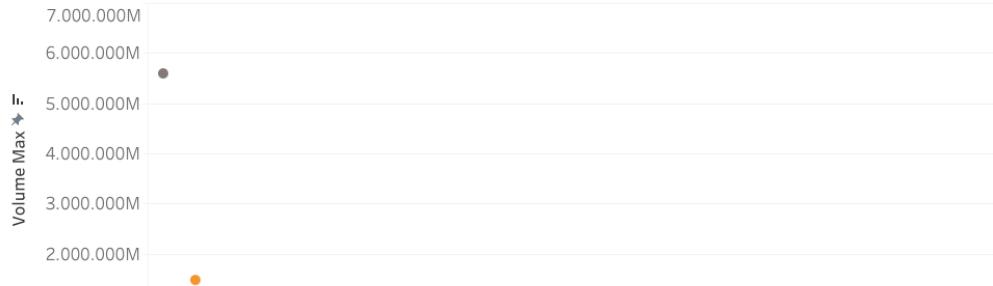
The financial challenge of the year

NOVARTIS DATATHON

online

Exploratory Data Analysis

Maximum volume per brand (only top brands)



Maximum volume per country



A wide range of orders of magnitude among countries and brands.





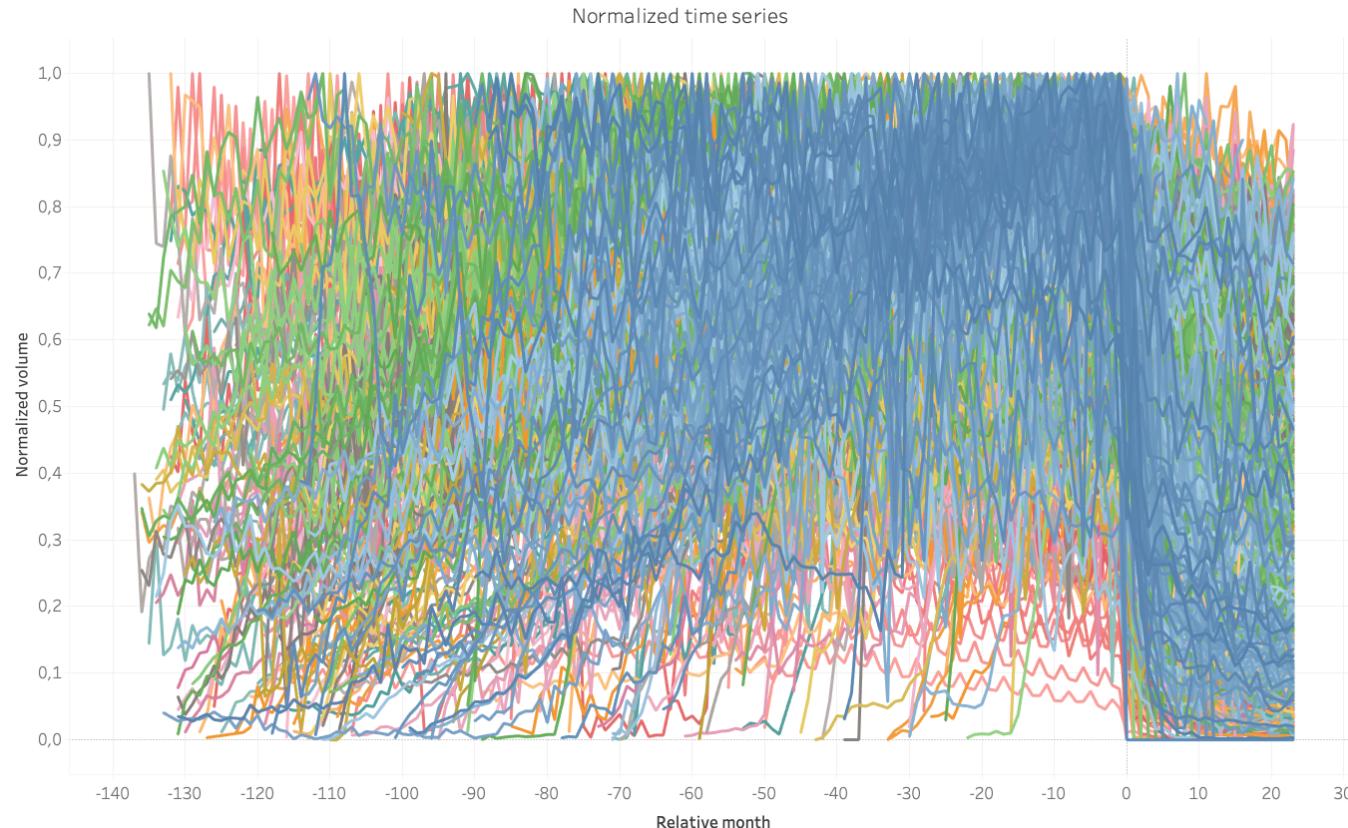
3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Exploratory Data Analysis





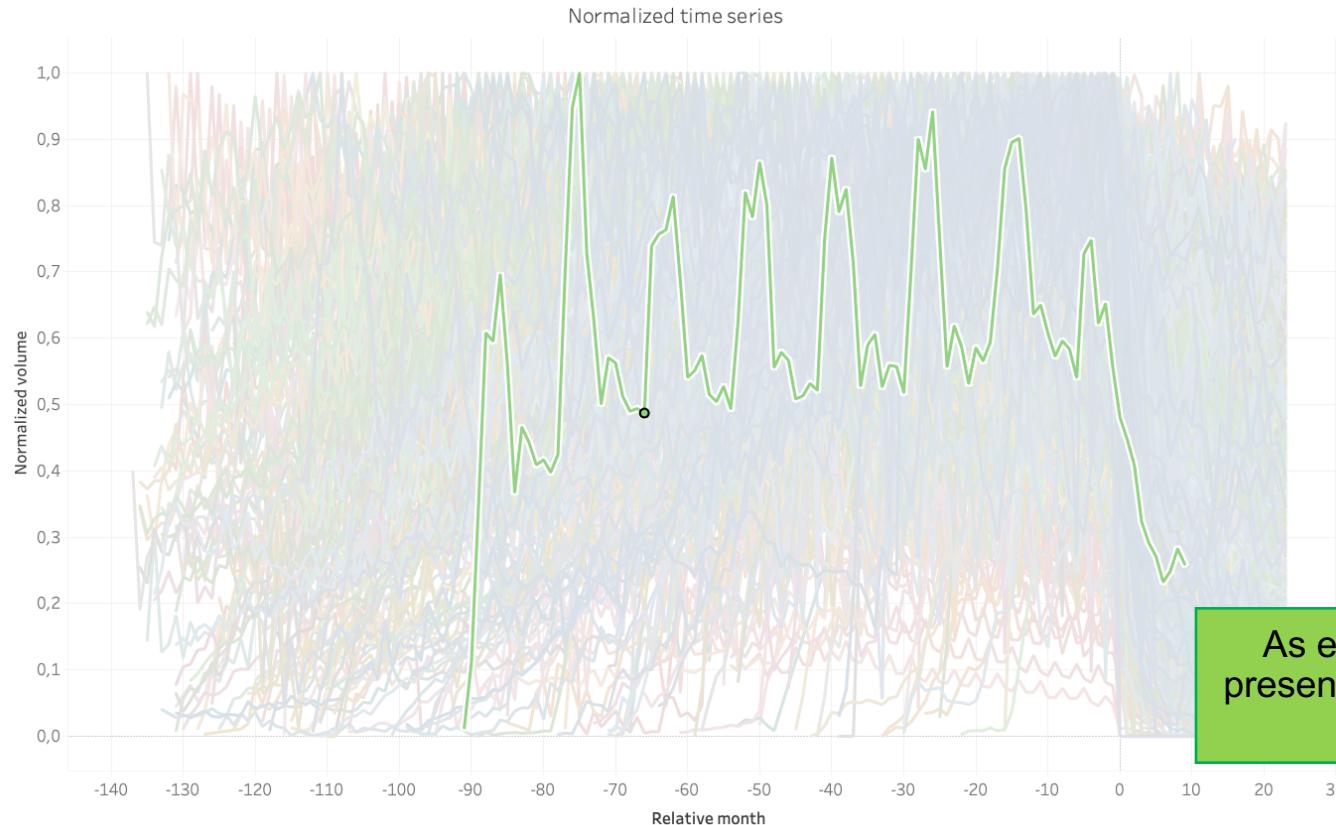
3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Exploratory Data Analysis





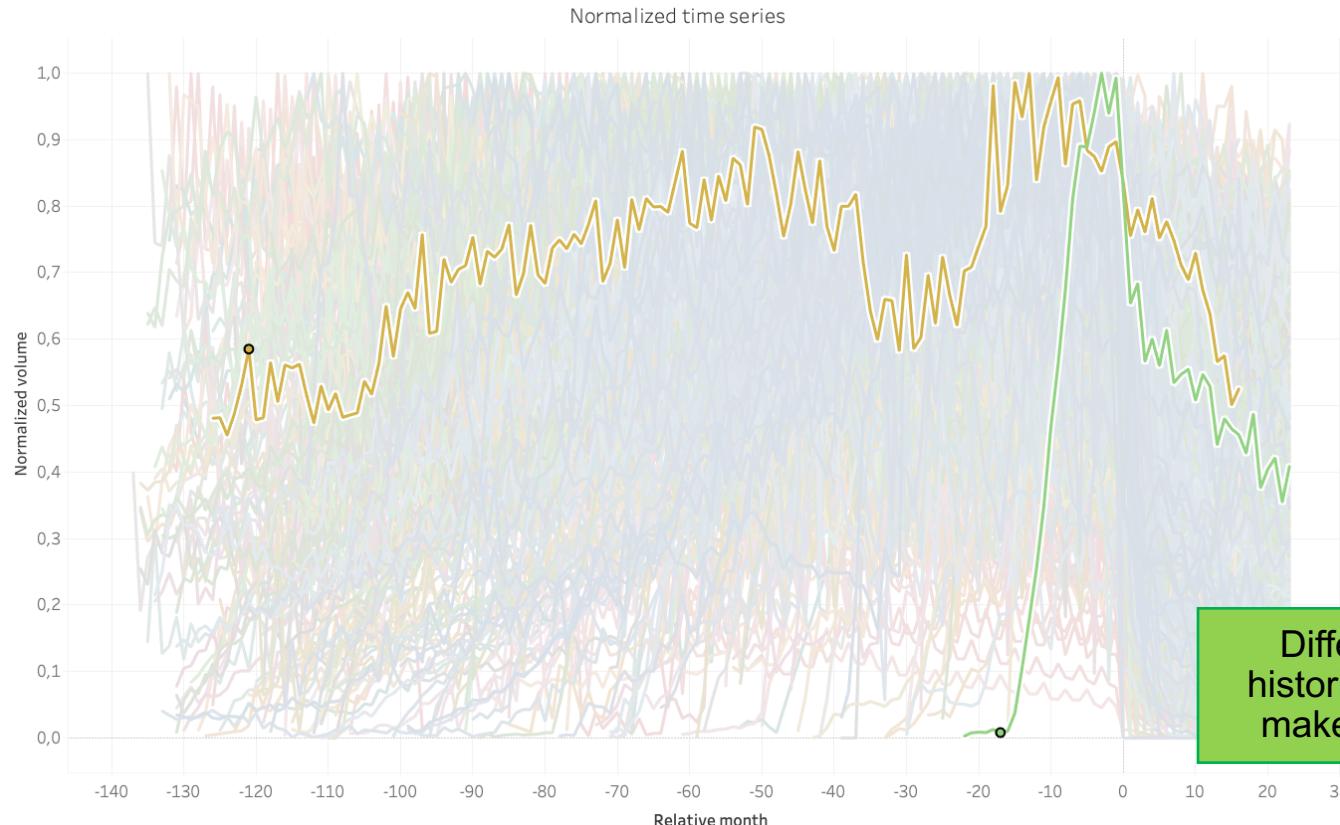
3rd edition

The financial challenge of the year

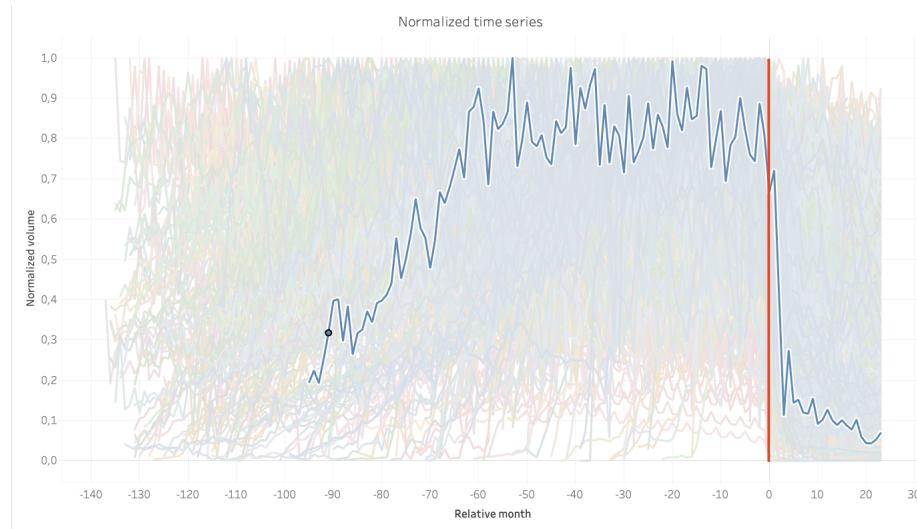
NOVARTIS DATATHON

online

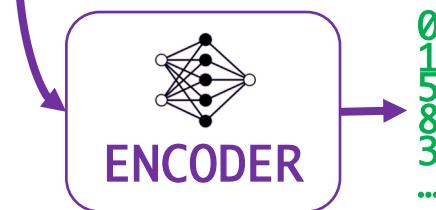
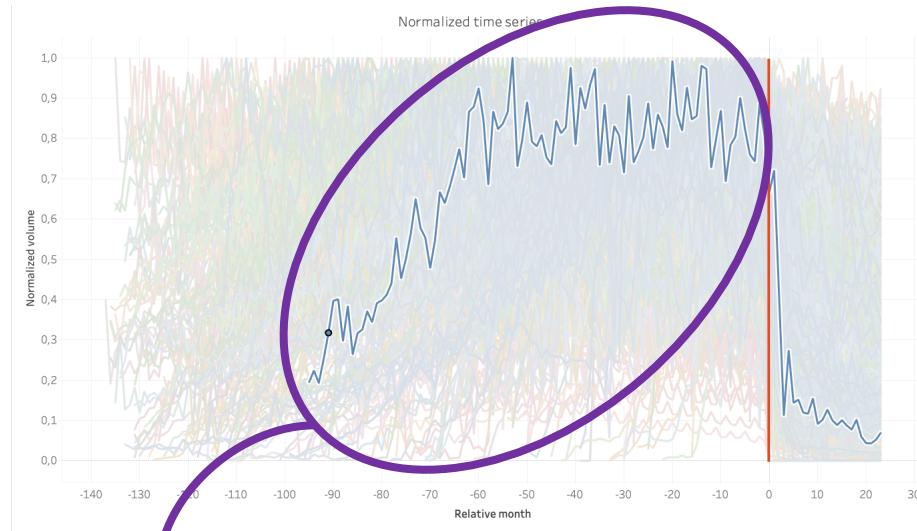
Exploratory Data Analysis



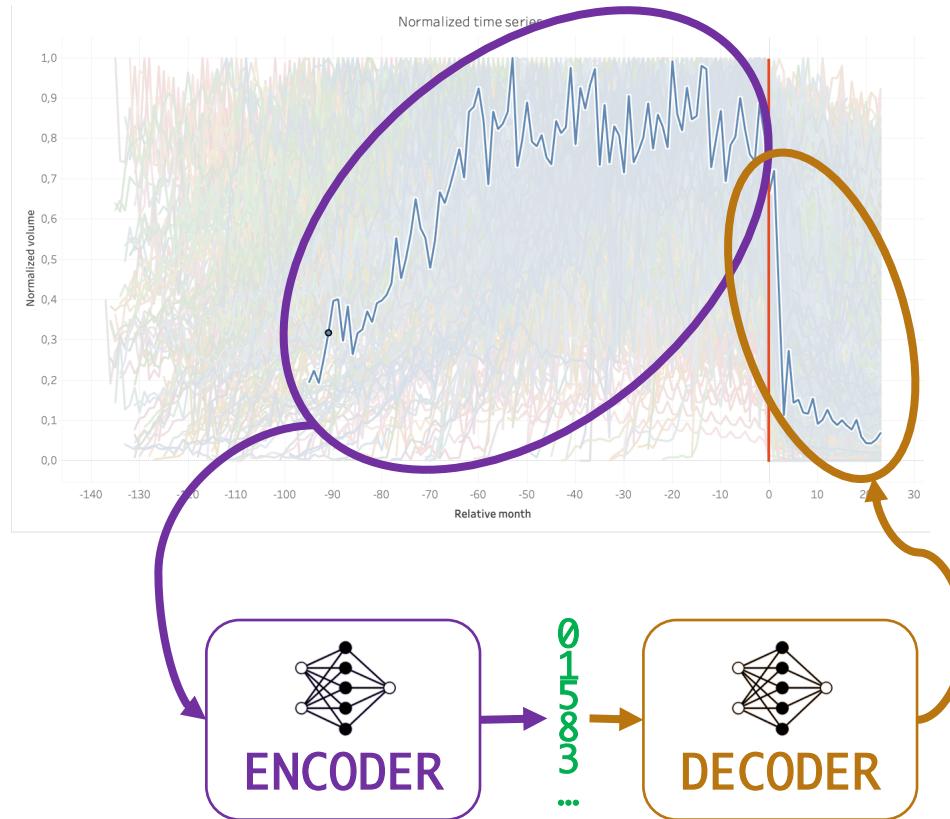
Model approach



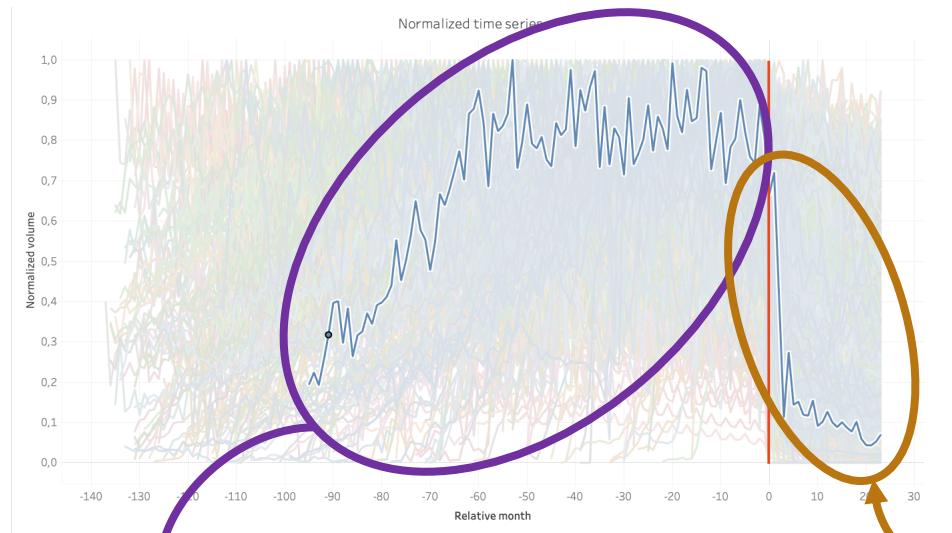
Model approach



Model approach

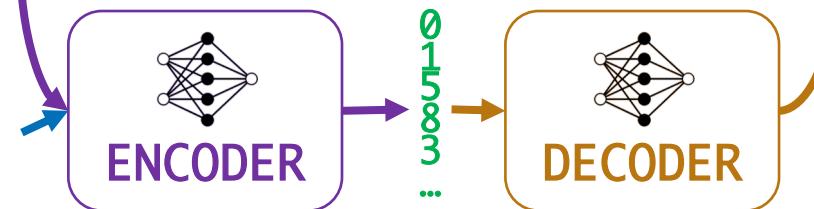


Model approach



Other features

Country (encoded)
Brand (encoded)
Package (encoded)
Therapeutical area (encoded)
Number of generics
Distribution channels
Month of the year (encoded)
...



Recurrent Neural Network Sequence to Sequence

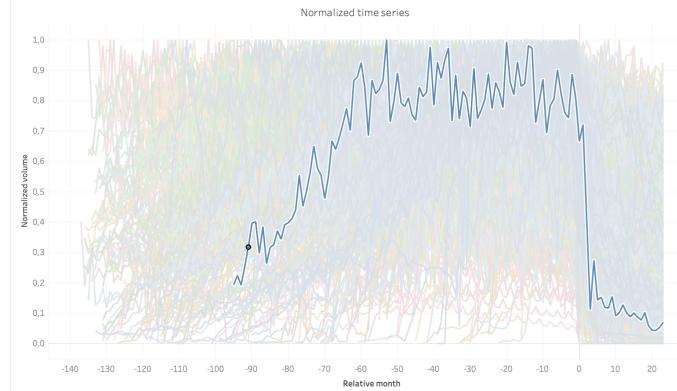
they are taking the → hobbits to Isengard

This book is largely concerned with Hobbits, and from its pages a reader may discover much of their character and a little of their history. Further information will also be found in the selection from the Red Book of Westmarch that has already been published, under the title of The Hobbit. That story was derived from the earlier chapters of the Red Book, composed by Bilbo himself, the first Hobbit to become famous in the world at large, and called by him There and Back Again, since they told of his journey into the East and his return: an adventure which later involved all the Hobbits in the great events of that Age that are here related.

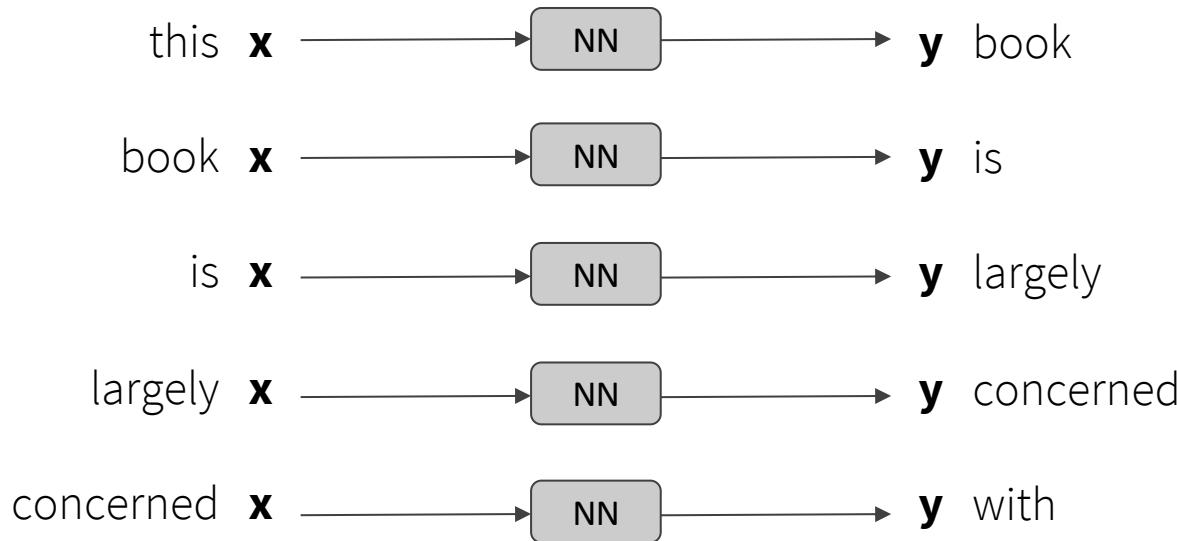
...

Recurrent Neural Network Sequence to Sequence

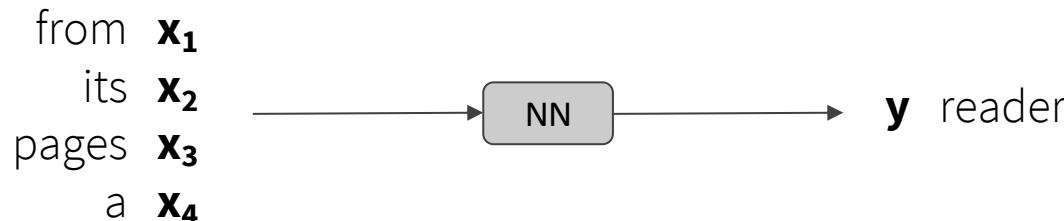
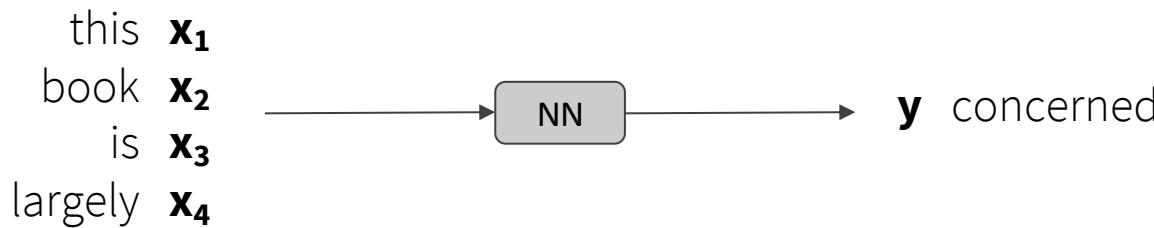
... 324.379 , 327.646 , 324.231 → 299.435 , 286.478 , 255.659 ...



Recurrent Neural Network Sequence to Sequence



Recurrent Neural Network Sequence to Sequence



Recurrent Neural Network Sequence to Sequence



3rd edition

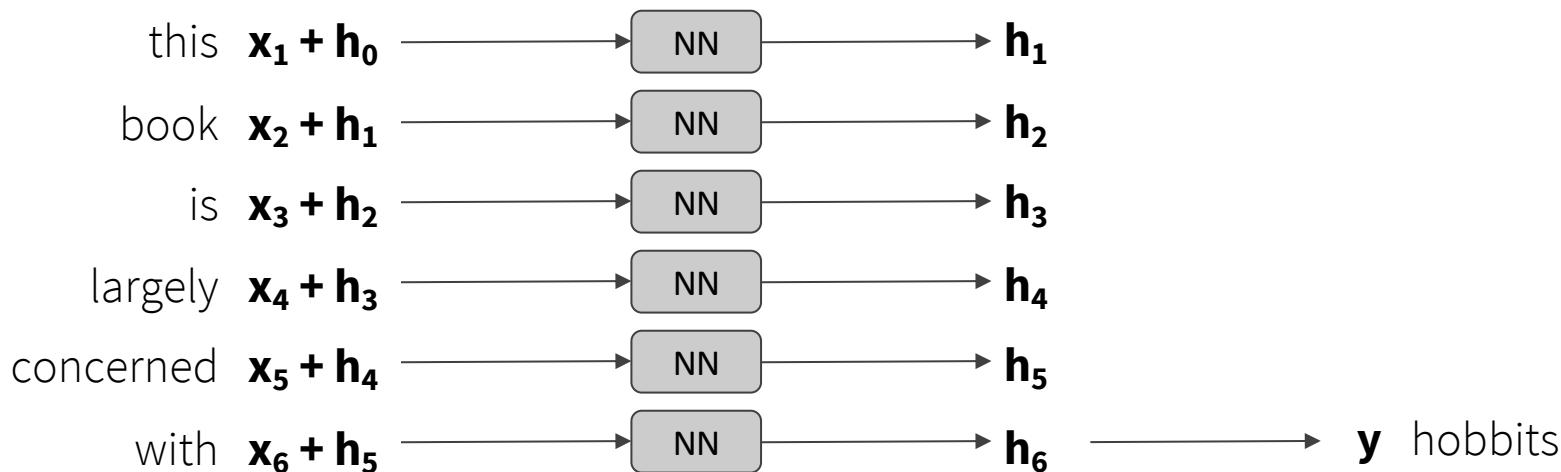
The financial challenge of the year

NOVARTISDATATHON

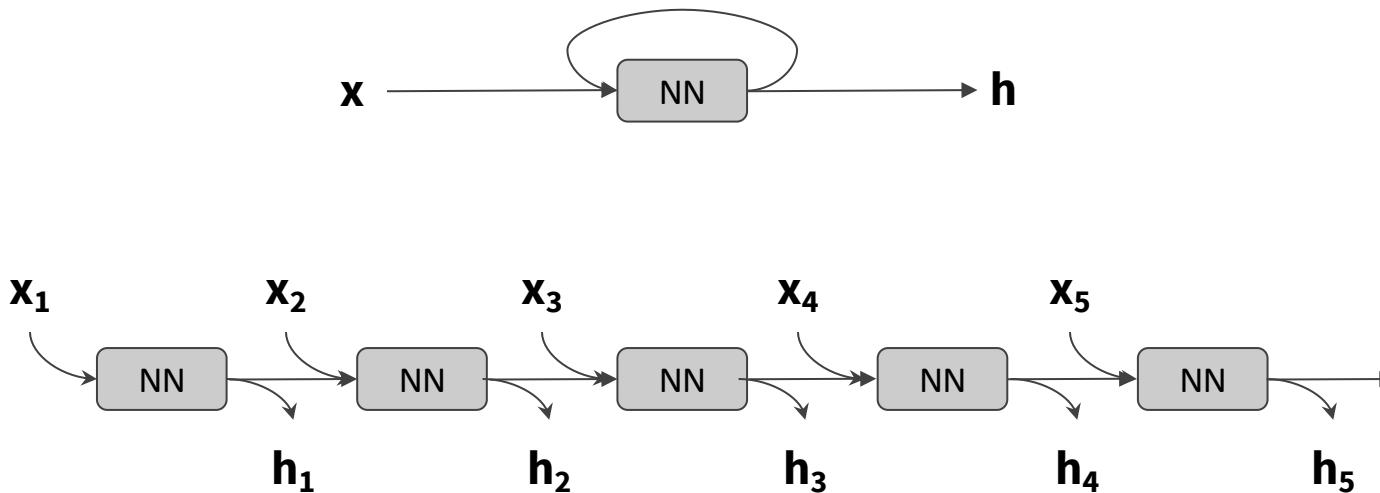
online



Recurrent Neural Network Sequence to Sequence



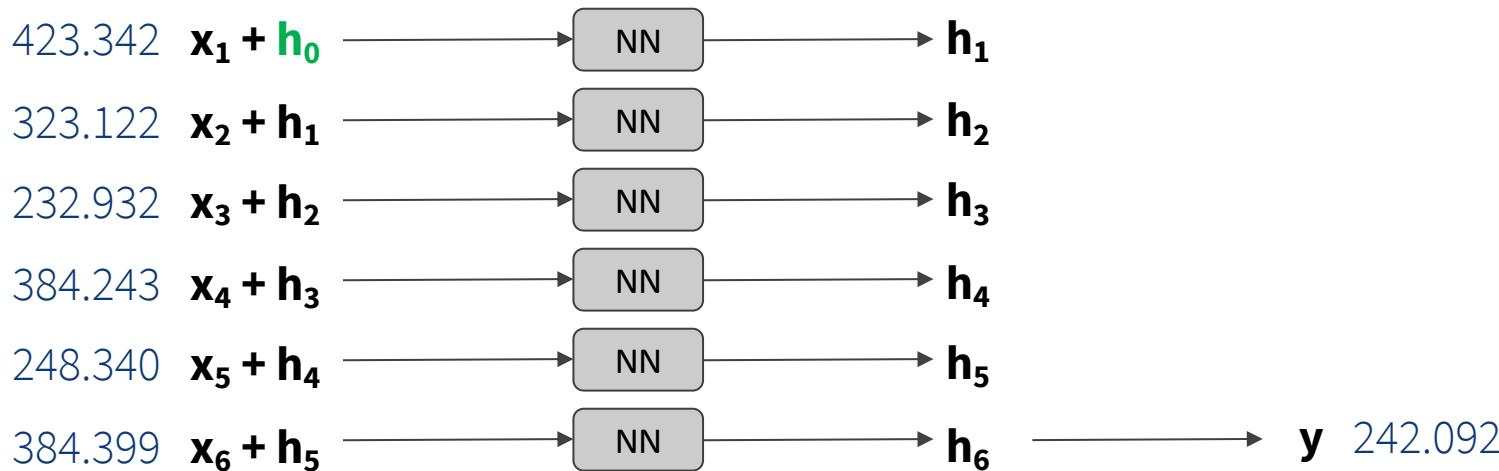
Recurrent Neural Network Sequence to Sequence

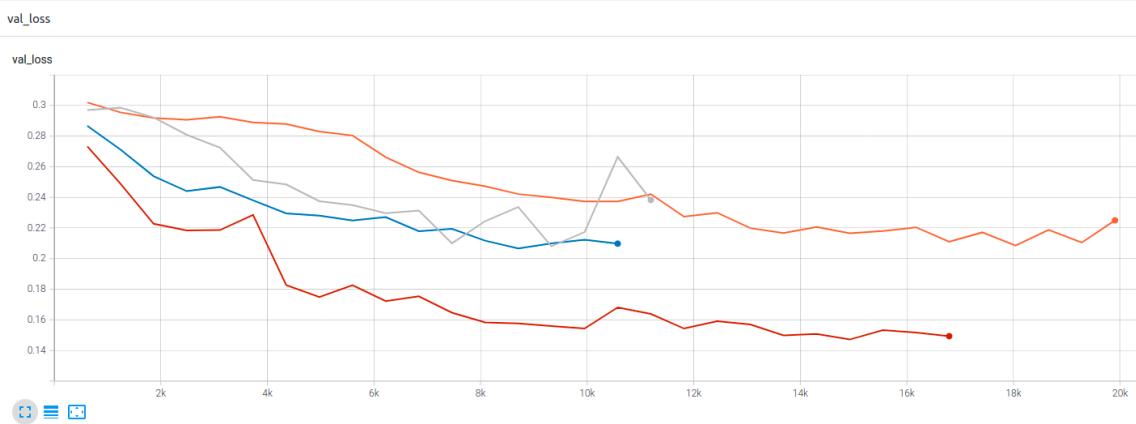


Recurrent Neural Network Sequence to Sequence

Historical values of
volumen sold per
drug

Other features: Number of generics,
therapeutical area, type of package, ...





- Basic encoder/decoder model ✓
- Custom metric ✓
- Data scaling ✓
- Using provided solutions ✓
- Scaling by country/brand ✓
- Include month (encoded) ✓
- Include rest of the provided features (encoded) ✓
- Include month in the decoder X
- Implement bigger batch size ✓
- Include confidence intervals in the predictions ✓
- Implement teacher forcing ✓
- Add dropout to embeddings ✓
- Further feature engineering
- Attention
- Blockchain
- ...

 PyTorch

 PyTorch
Lightning



TensorBoard

 NOVARTIS



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Learning Rates

- More feature engineering.
- More parallel work.
- More previous preparation.





3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Winners

- Random Forest
- 1 model per month to predict (24)
- Predicted output of month m used as an input for the $m+1$ prediction
- Confidence intervals calculated using cross-validation



3rd edition

The financial challenge of the year

NOVARTISDATATHON

online

Winners

- Ensamble of ML models
- Hyper-parameter optimization
(+10k executions)



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Winners

- GLM and XGBoost
- Momentum features (e.g. growth prior to generics entry)



3rd edition

The financial challenge of the year

NOVARTIS DATATHON

online

Moral winners





3rd edition

The **financial challenge** of the year

NOVARTIS DATATHON

online

#godatathon