DeepPatient: Leveraging EHR data to predict patient outcomes

Tiago T. V. Vinhoza

Big Data Analytics and Data Science Department

DOCNET

DSPT
DATA SCIENCE PORTUGAL

Digitized Patient

DOCTOR IS IN

Google is using 46 billion data points to predict the medical outcomes of hospital patients

Digitized Patient - DOCNET

291 million datapoints

Predicting Patient Deterioration

- Early Detection of patient deterioration
  - Improve clinical outcomes:
    - For each hour sepsis treatment is delayed, the patient's risk of death increases by 4 percent. (New England Journal of Medicine).
  - Reduce hospital costs:
    - A patient at the ICU costs 1000 euros a day to the hospital.

DSPT
DATA SCIENCE PORTUGAL

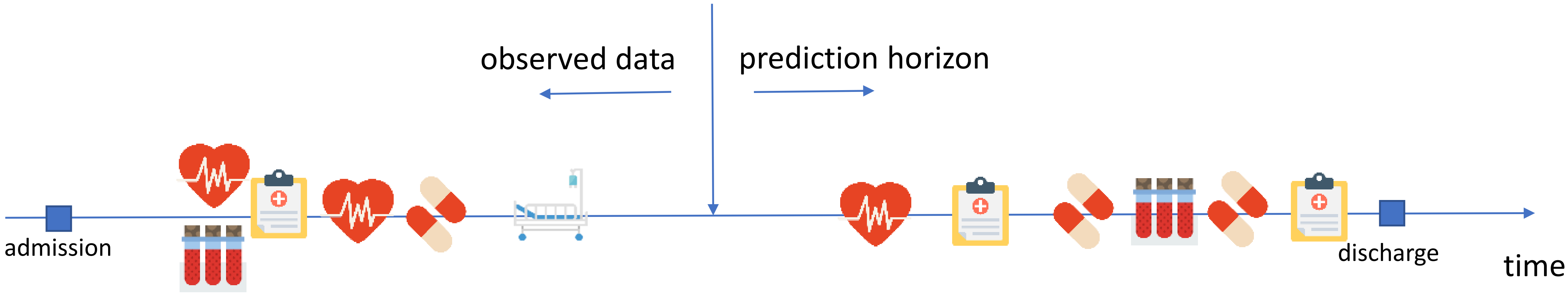
Predicting Patient Deterioration

- Example: NEWS2 (National Early Warning Score – UK – only vital signs

Physiological parameter	Score 3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

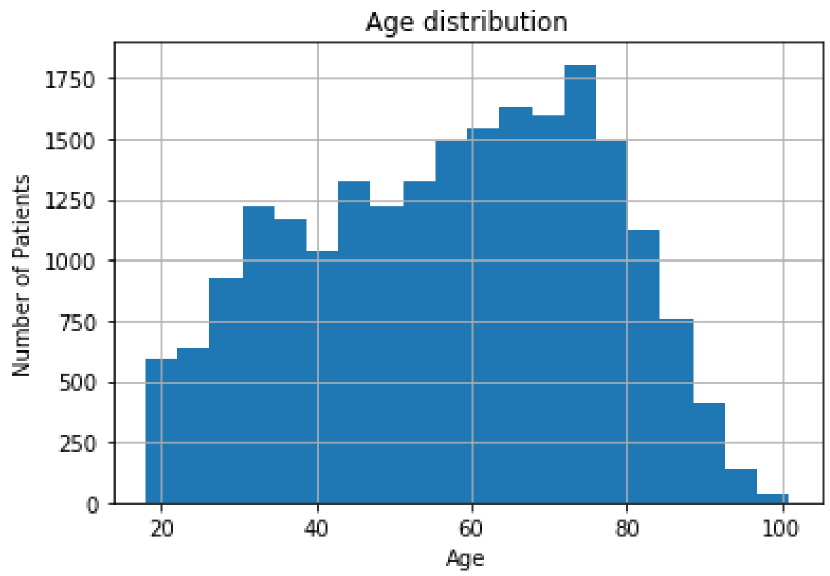
Some definitions before looking at the data

- An **episode** is a sequence of events that describe the stay of a patient in the hospital.
- A **snapshot** is a picture taken from an episode at a given point in time.



Exploratory Data Analysis: Demographics

- Training set: 21470 episodes (1 snapshot / day) from 1-year period.
  - Excluded cases: pediatric patients (<18), patients under paliative care.
  - Average Age: 57 years old
  - Sex: 44% men, 56% women

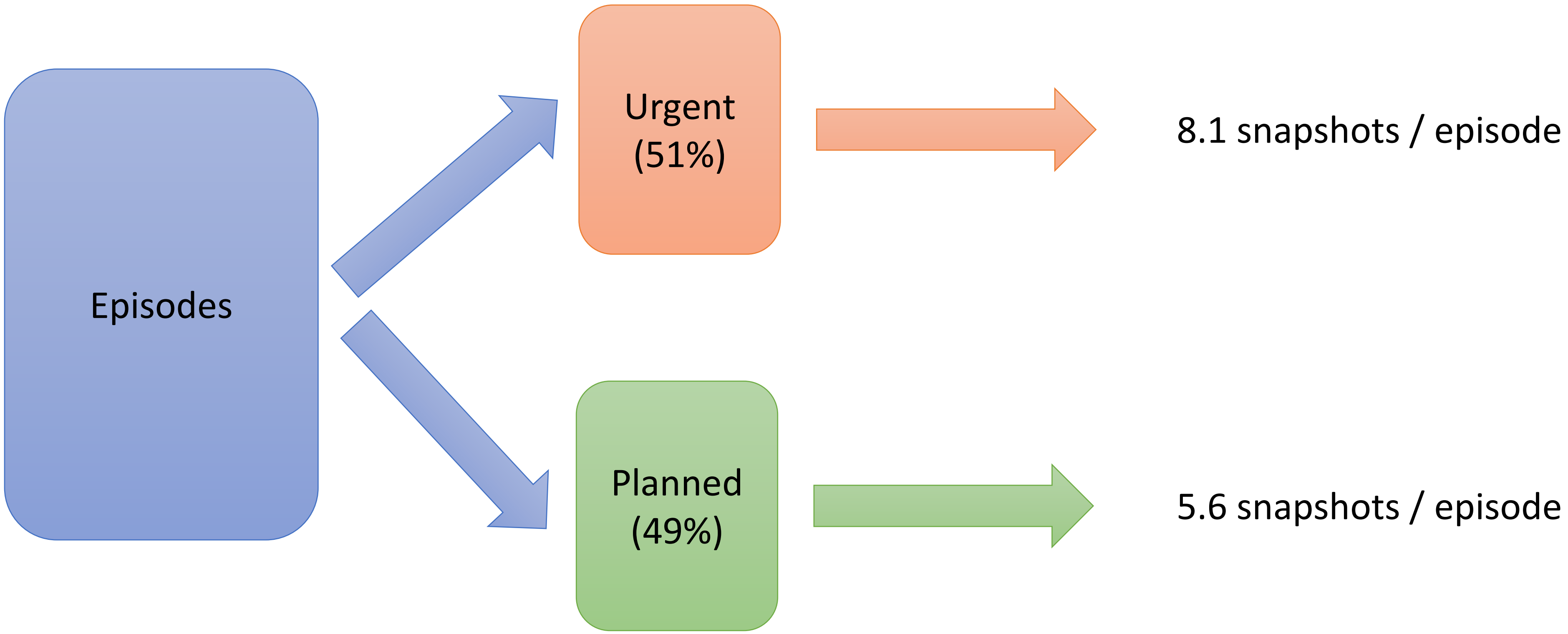


- Total number of snapshots: 147676

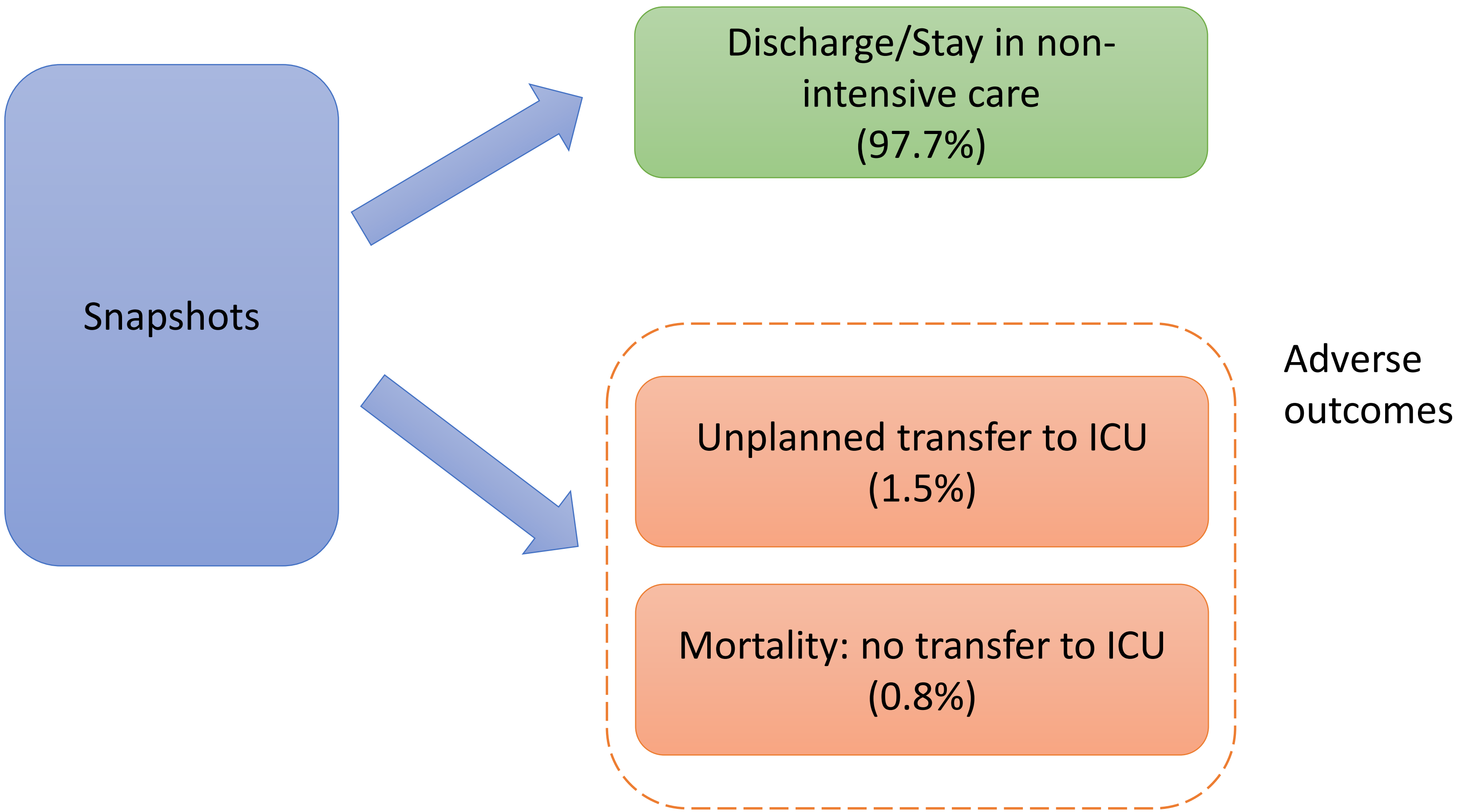
6.9 snapshots / episode
(approximate length of stay)

DSPT
DATA SCIENCE PORTUGAL

Exploratory Data Analysis: Admission



Exploratory Data Analysis: 7-day Outcome



Major Challenge in Healthcare

Different units emphasize different signals when monitoring the patient

- Heart Rate, Blood Pressure
  - Cardiology: ~4x a day, Plastic Surgery: ~1x day
- Creatinine
  - Renal Transplant: ~1x a day, Internal Medicine: once every 2.5 days

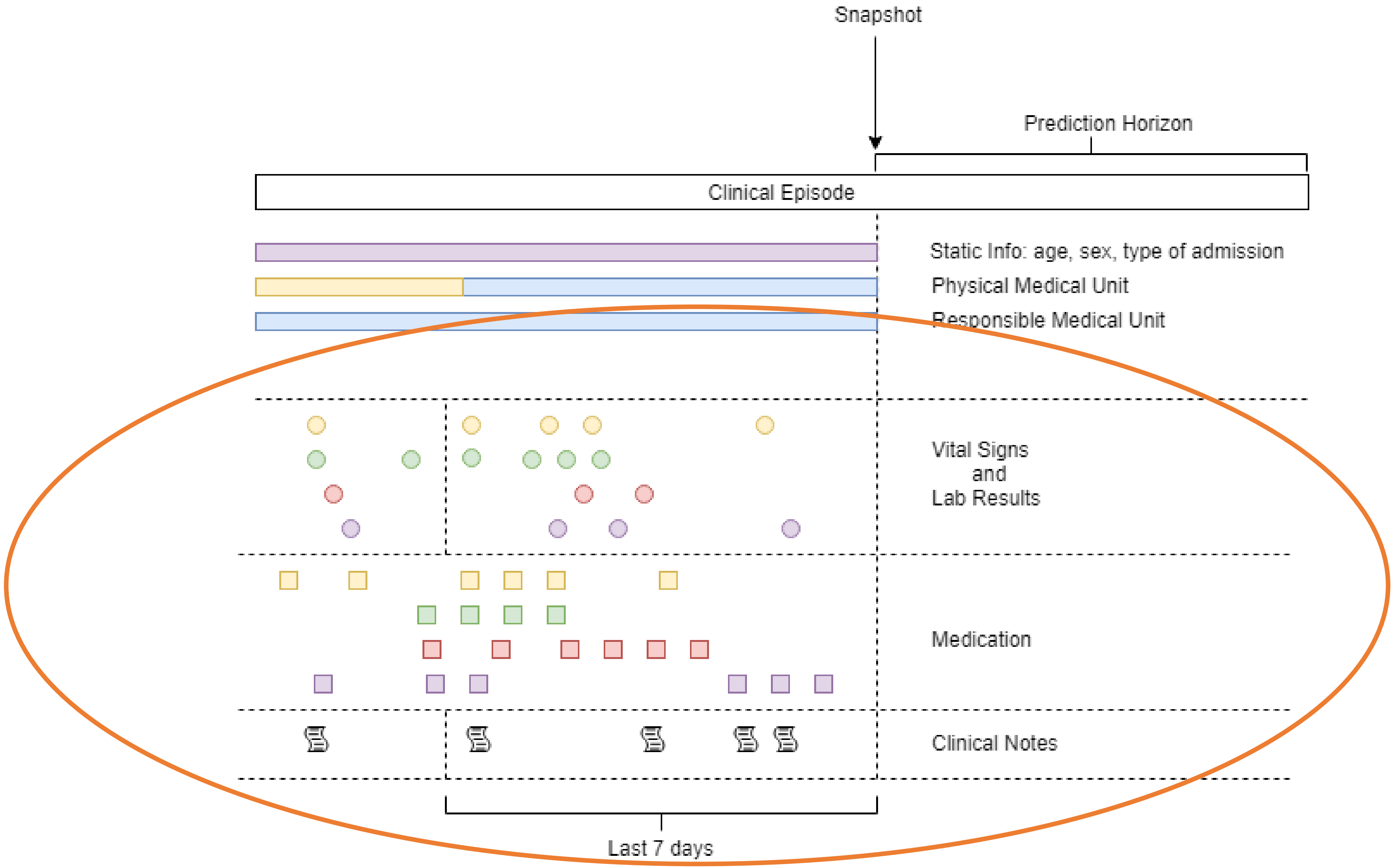
Few measurements means more missing values

↓

Concept of Informative Missingness

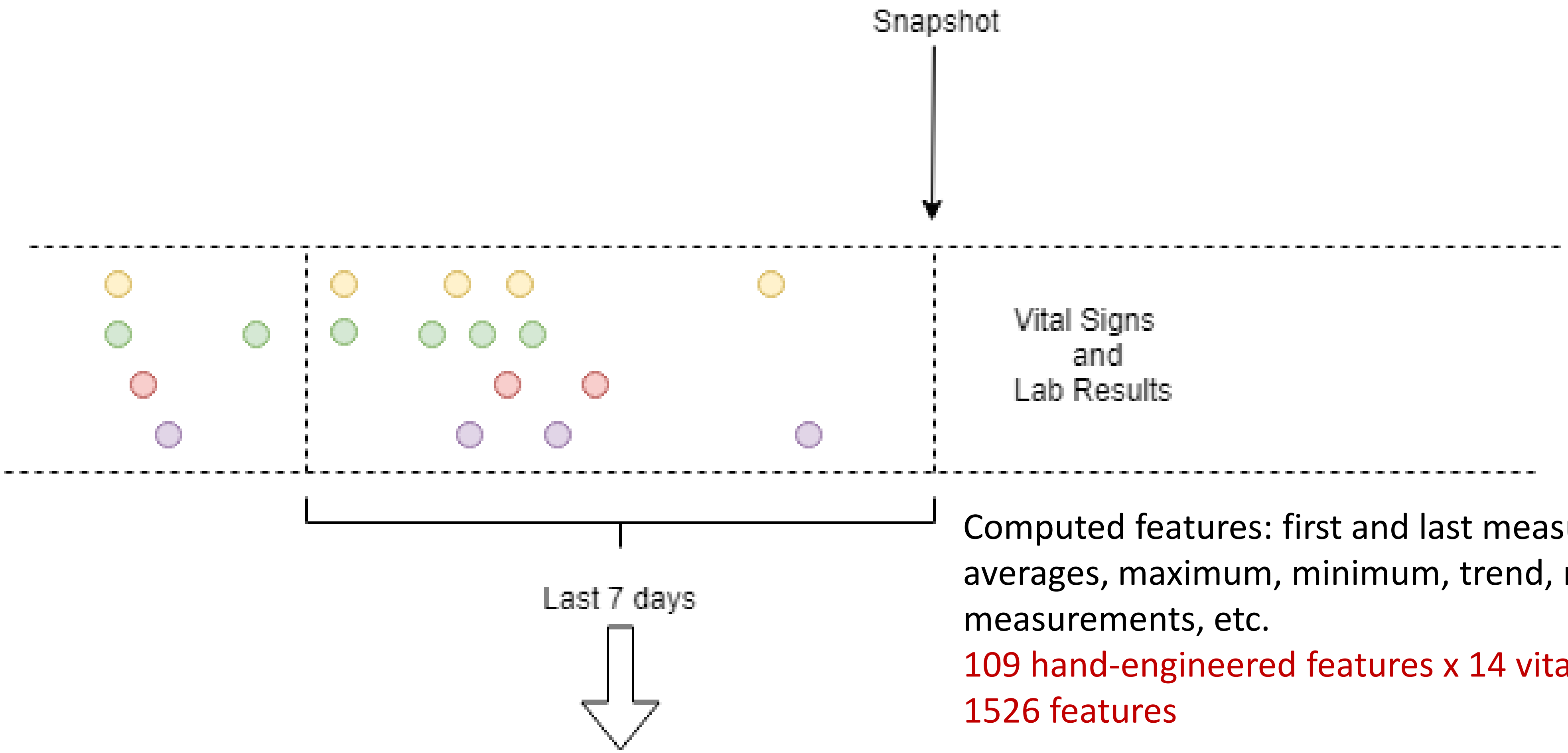
DSPT
DATA SCIENCE PORTUGAL

Patient Representation



Non-uniformly spaced time-series.
How to represent?

Vitals + Labs

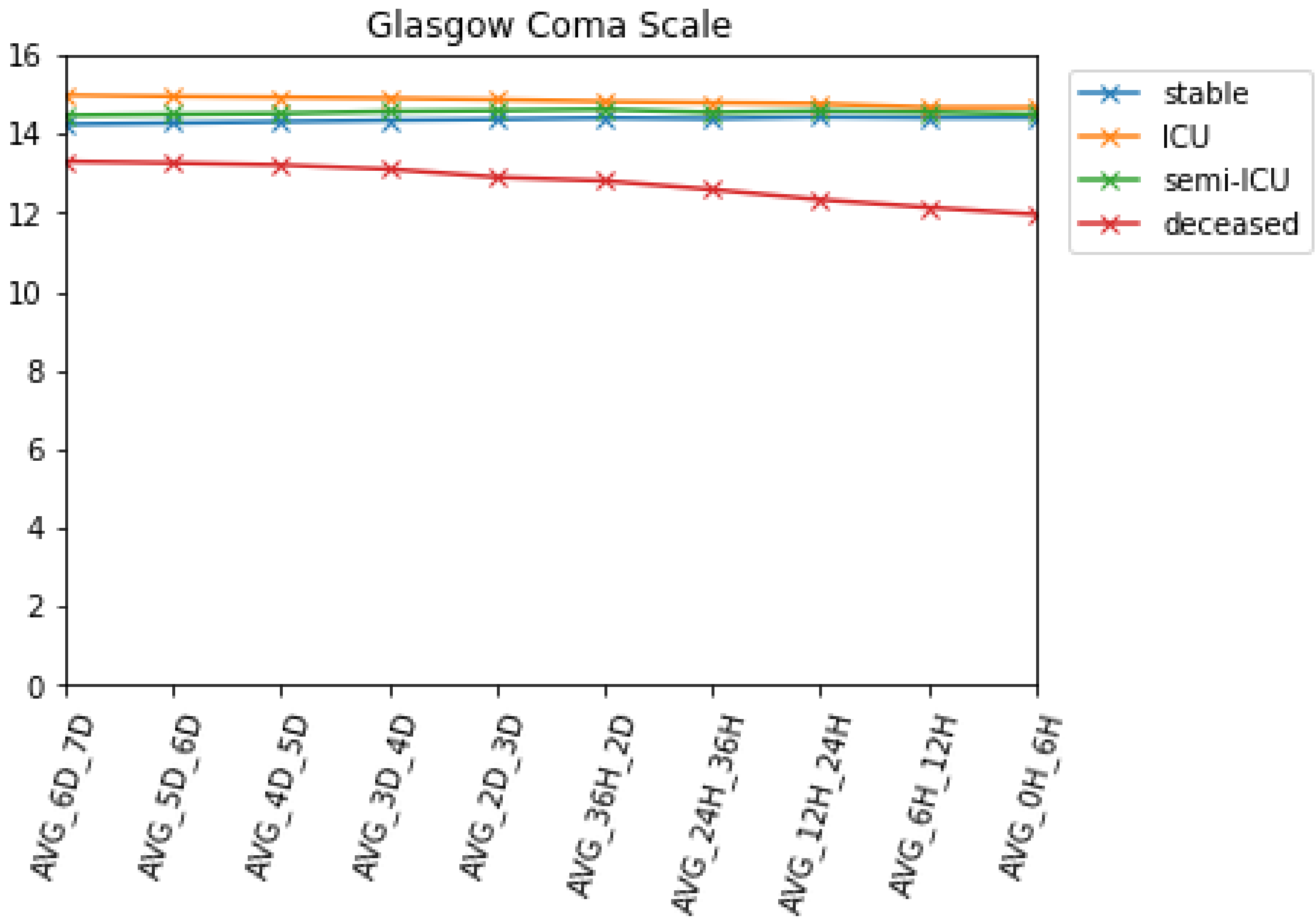


Computed features: first and last measurements, averages, maximum, minimum, trend, number of measurements, etc.

109 hand-engineered features x 14 vitals/labs = 1526 features

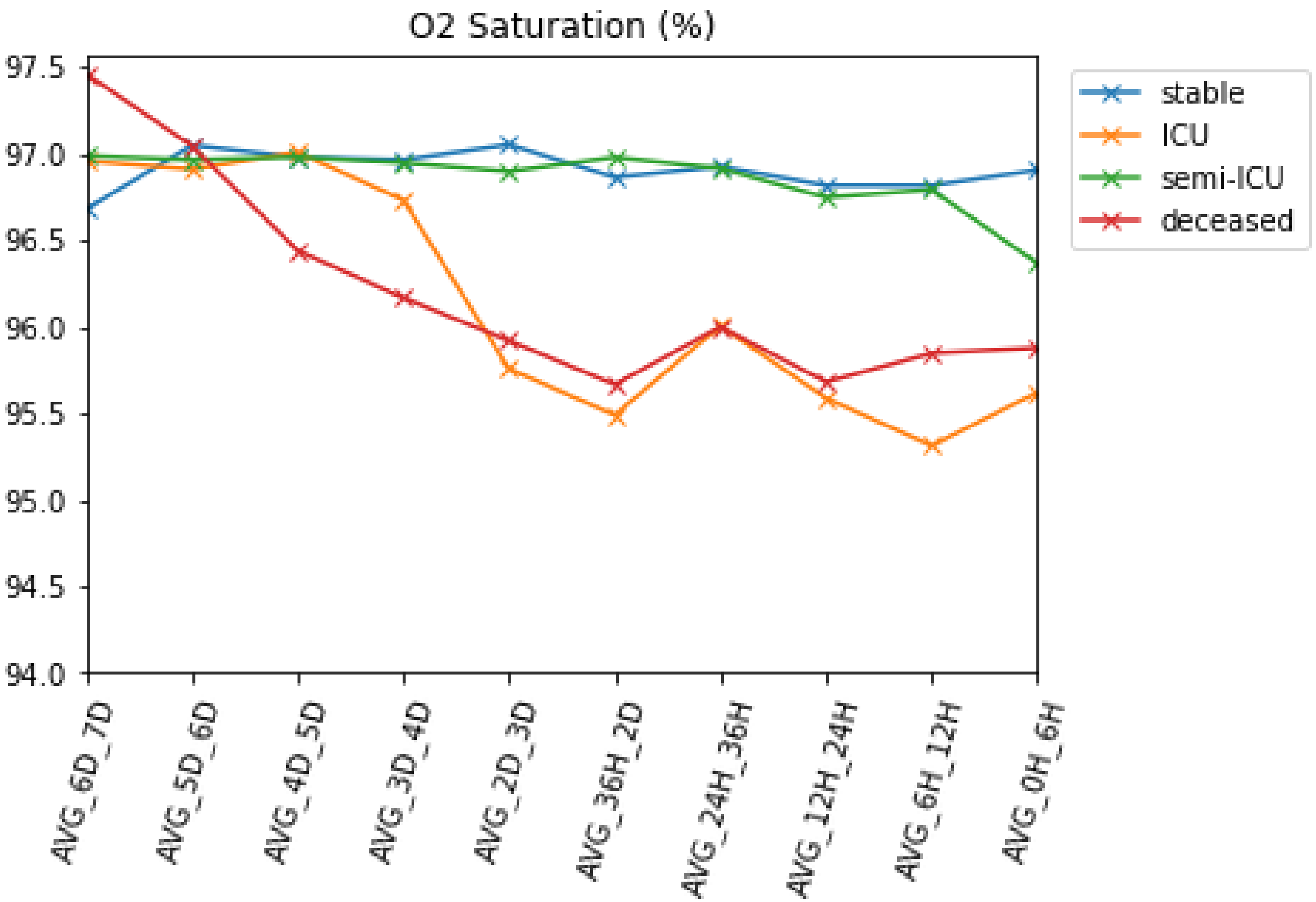
Col 1	Col 2	...	Col 109	Col 1	Col 2	...	Col 109	Col 1	Col 2	...	Col 109	Col 1	Col 2	...	Col 109
37.2	36.9	...	36.6	8.50	NULL	...	1.23	12	NULL	...	13	120	132	...	NULL

Exploratory Data Analysis



- All non-surgical units
- Steady decline of consciouness level for the deceased patients

Exploratory Data Analysis

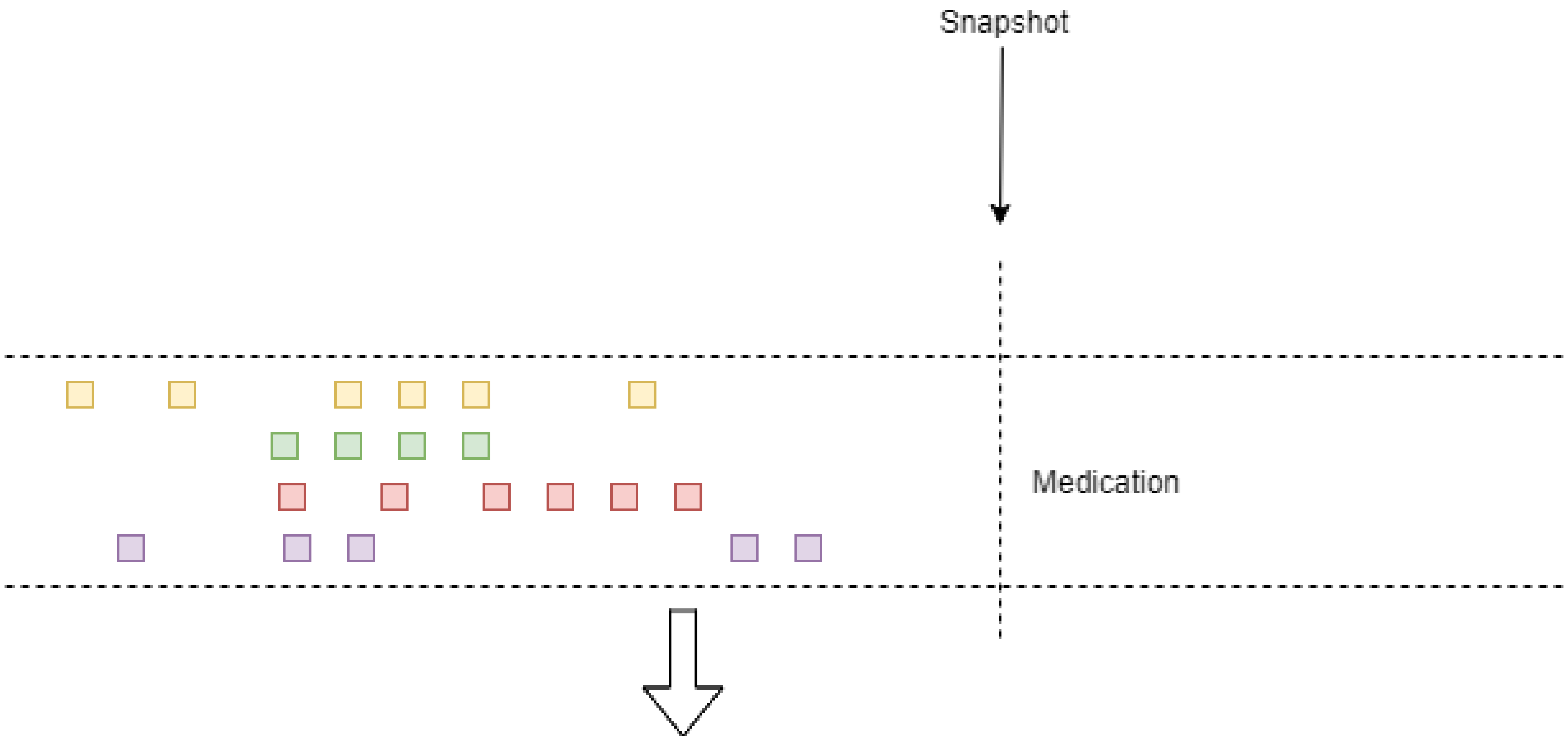


- General Surgery Unit
- 2 percentage point drop for ICU-bound and deceased patients

Medication

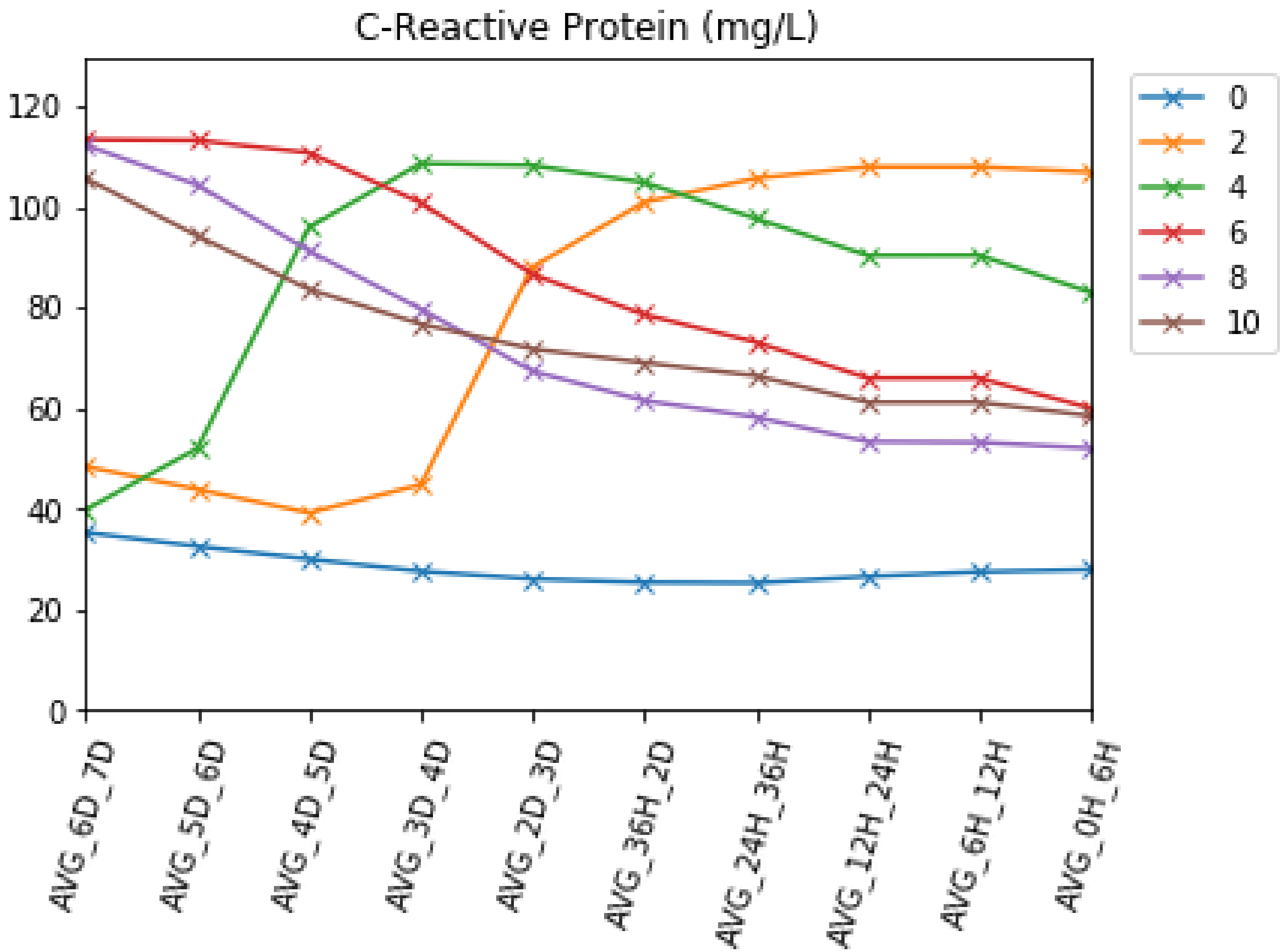
- Very large number of medication codes
  - Group by general role:
    - Anti-Infection – Antibacterial
    - Cardiovascular – Anti-Arrithmya
- Feature Engineering
  - For each medication compute 4 features:
    - number of consecutive days taking medication.
    - total number of days taking medication.
    - days since last intake.
    - took medication on current day? (yes/no)
  - 4 computed features x 93 medication groups = 372 features

Medication



str	acc	days_since_last	took_curr day	str	acc	days_since_last	took_curr day	str	acc	days_since_last	took_curr day	str	acc	days_since_last	took_curr day
1	6	1	0	0	4	2	0	1	6	0	0	1	5	0	1

Exploratory Data Analysis



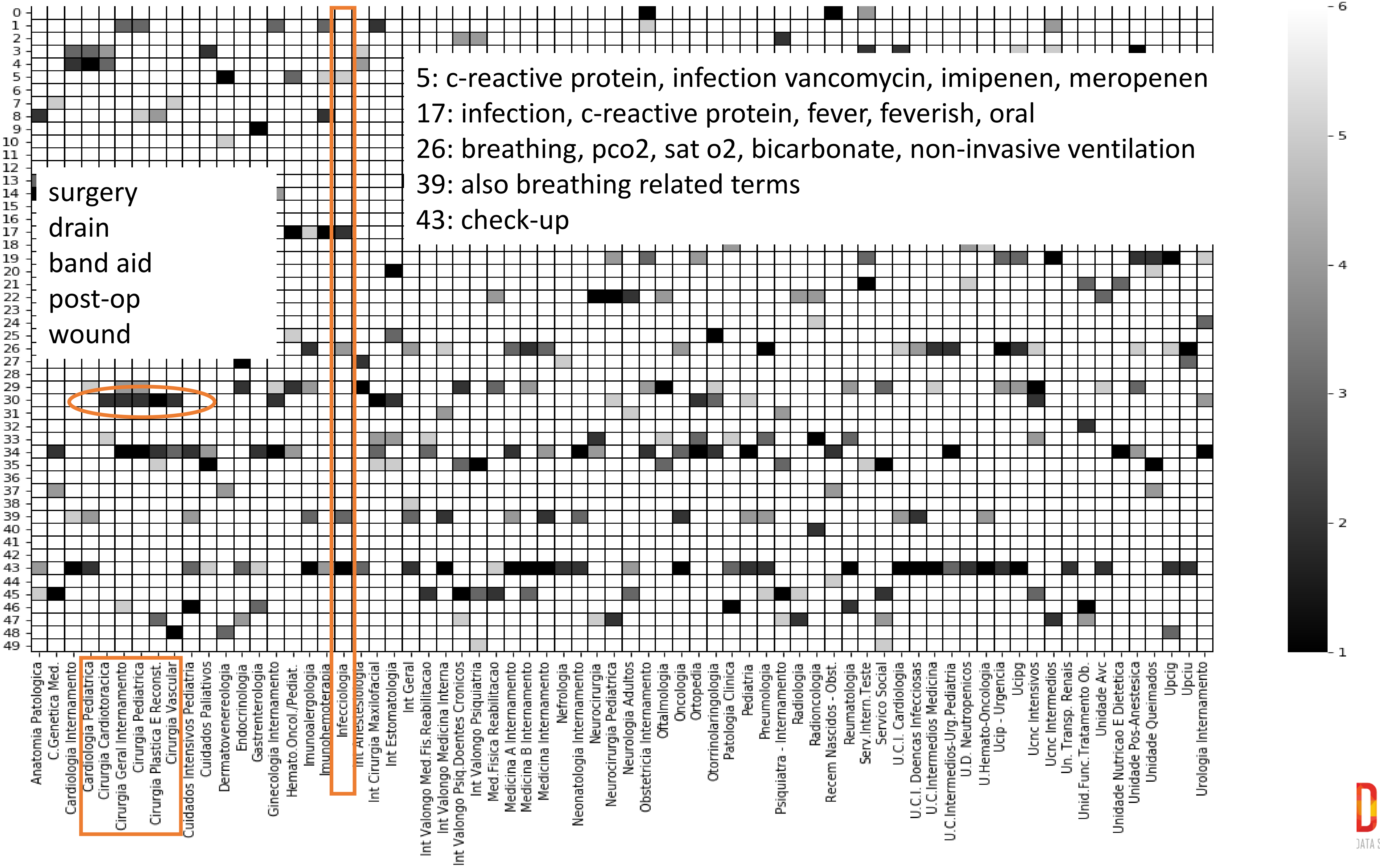
- Internal Medicine Unit
- Consecutive days taking antibacterials
- It takes on average 2 days to reduce C-reactive protein levels (proxy for infection).

DSPT
DATA SCIENCE PORTUGAL

Clinical Notes

"The gold is in the clinical notes" - Edward Choi, co-author of the study titled Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

- Topic Modeling using Latent Dirichlet Allocation:
    - Noisy data (typos, names, overhead).
    - Different ways to write notes in different hospital units.


5: c-reactive protein, infection vancomycin, imipenen, meropenen
17: infection, c-reactive protein, fever, feverish, oral
26: breathing, pco2, sat o2, bicarbonate, non-invasive ventilation
39: also breathing related terms
43: check-up
surgery
drain
band aid
post-op
wound

Preliminary Results

- Baseline Models:

Emmanuel Ameisen
AI Lead at Insight AI @EmmanuelAmeisen
Mar 6 · 9 min read

Always start with a stupid model, no exceptions.

- LSTM:

Eric Topol
@EricTopol
Following

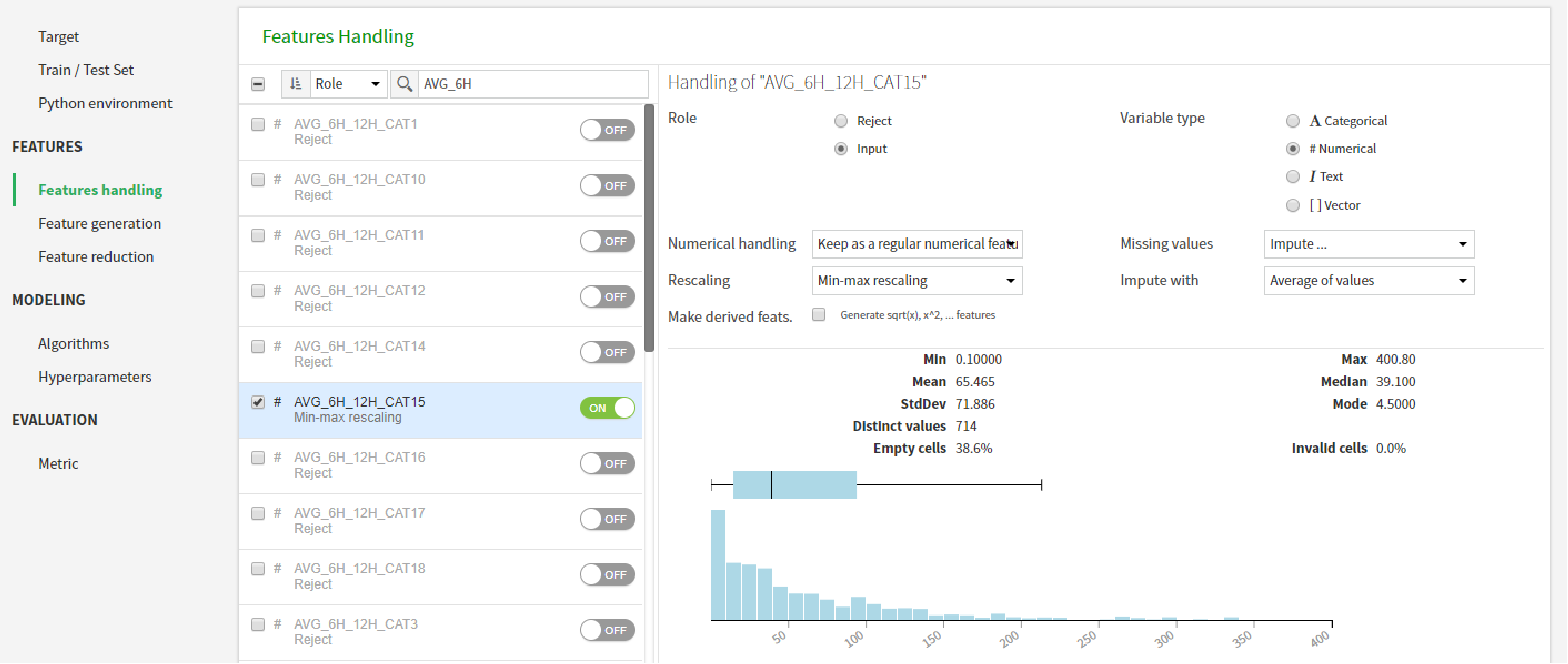
We still use dumb algorithms (rules-based, heuristic, univariate) in medicine, developed decades ago. Eagerly await validated smart ones w/ deep neural networks #AI

Preliminary Results (Dataiku)

- Baseline Models on Dataiku
  - Random Forests
  - Gradient Tree Boosting
  - XGBoost

- Different Feature Sets
  - Vitals/Labs
  - Vitals/Labs + Medication

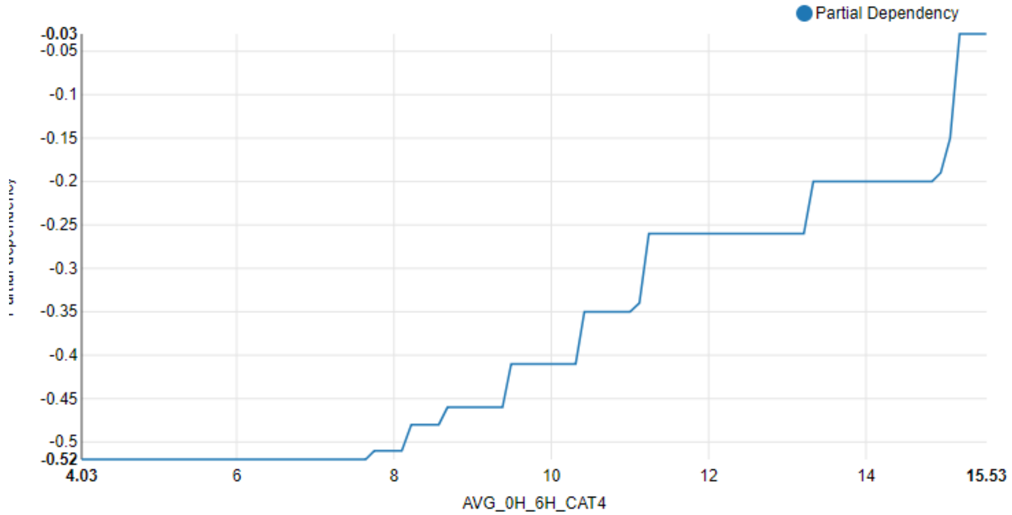
Preliminary Results (Dataiku)



DSPT
DATA SCIENCE PORTUGAL

Preliminary Results (Dataiku)

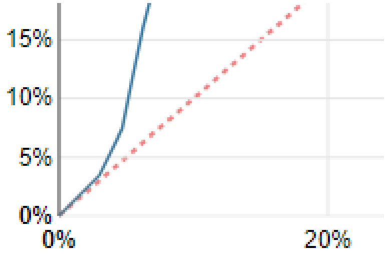
- Vitals – Gradient Tree Boosting AUC: 0.735
  - Precision 10%, Recall 26% (max F1-score – 5.6% rate of alarm)
  - Precision 20%, Recall 9% (about 1% rate of alarm)
  - Recent measurement were the most important



Preliminary Results (Dataiku)

- Vitals+Medication – XG Boost AUC: 0.781
  - Precision 23%, Recall 11% (about 1.1% rate of alarm)

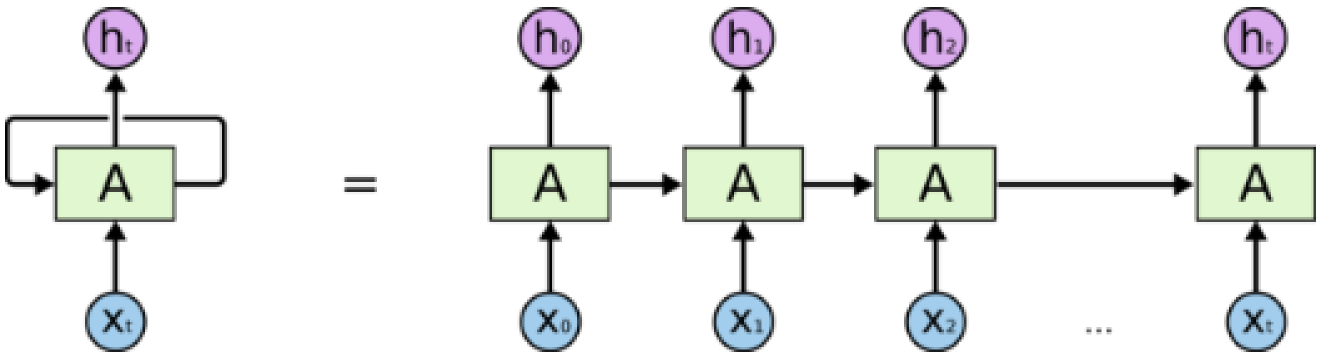
- Random Forest: AUC 0.73 (bad ROC shape)



Deep Patient approach

Preliminary Results (LSTM)

- Powerful to learn from sequence data.
- Capture long term dependencies and non-linear dynamics.



Preliminary Results (LSTM)

- Features
  - Responsible unit (one-hot encoded).
  - Vitals/Labs: taking AVG_24H features (make compatible with timesteps).
- Missing values
  - Forward filling + imputation with average.
- Normalization: [0,1] interval.
- Network starts deciding with 3 days of data.

DSPT
DATA SCIENCE PORTUGAL

Preliminary Results (LSTM)

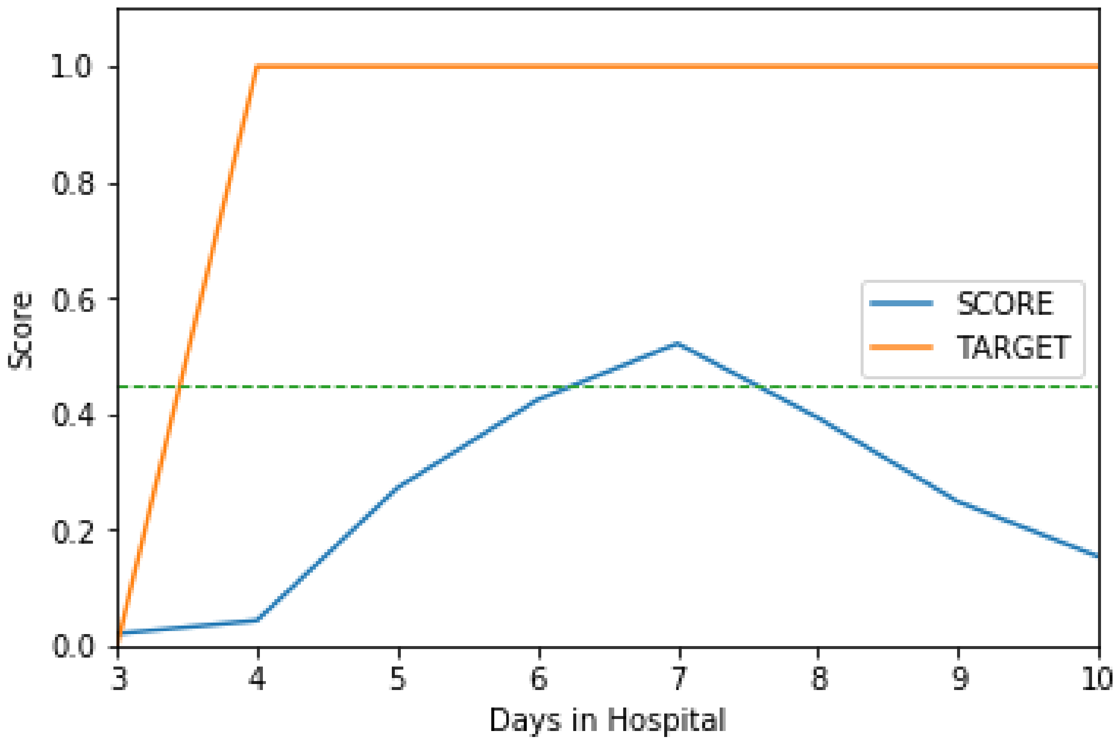
- Two LSTM layers (64 cells) + dense layer (sigmoid).
- 7 time-steps (use last week of data as input).
- 15% of training data used for validation.
- Trained for 40 epochs – used best epoch based on validation error.

Preliminary Results (LSTM)

- AUC: 0.745
  - Precision: 20%, Recall: 28% (rate of alarm – 3.5%)
  - Precision: 24%, Recall: 10% (rate of alarm – 1%)

- Scoring an episode:

(patient goes to ICU on day 11,
alarm sounds on day 7)



But how to find the best model using 291 million datapoints?

Machine Learning?

26.000 Models tested/day

Credits:

Rui Quintino

devscope

DSPT
DATA SCIENCE PORTUGAL

Conclusions

- Electronic Health Records: rich source of data, first step towards personalized medicine.

- Challenges:
  - Inclusion/Exclusion criteria.
  - Dealing with missing values – treat missingness as features.
  - Best way to incorporate clinical notes.
  - Hyperparameter tuning.
  - LSTM Model Interpretation – attention mechanism.