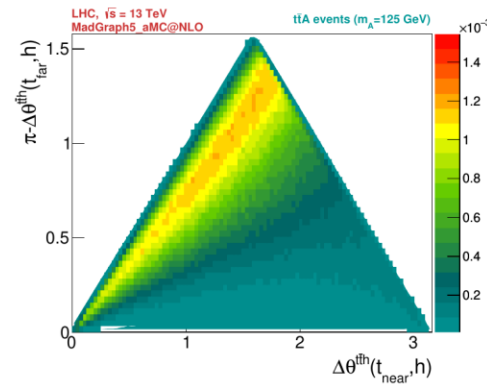
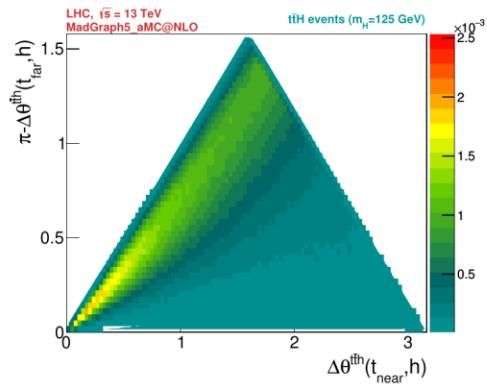
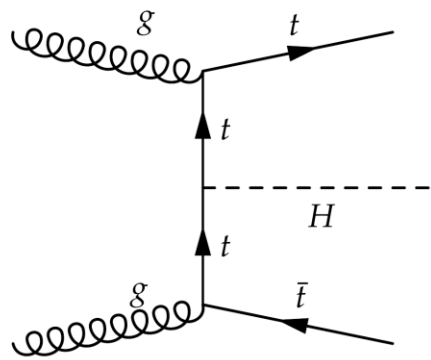
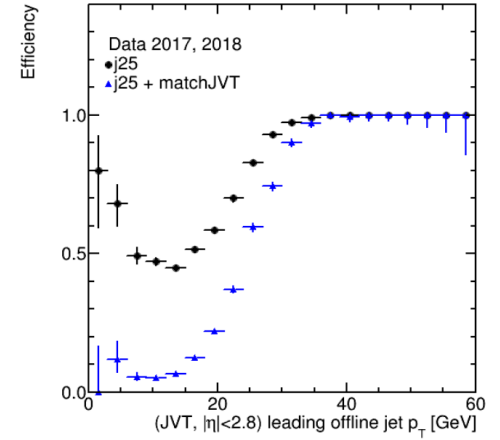
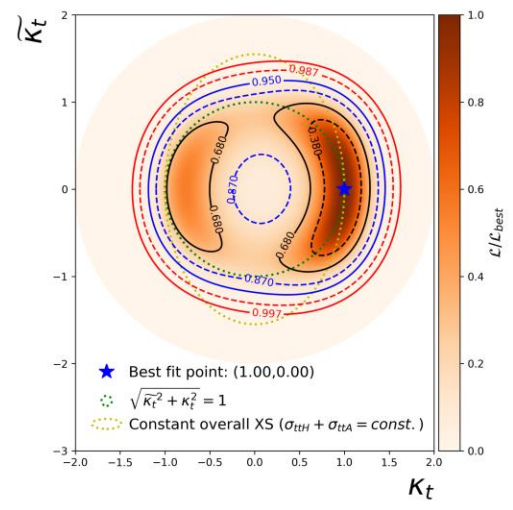
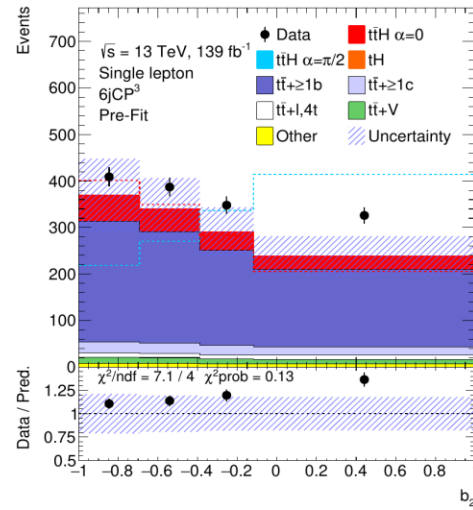


Data science in collaborative innovation projects

Emanuel Gouveia
DTx CoLab

Data Science in Fundamental Physics
and the bridge to industry & society
IGFAE, Santiago de Compostela
07/06/2024

My background





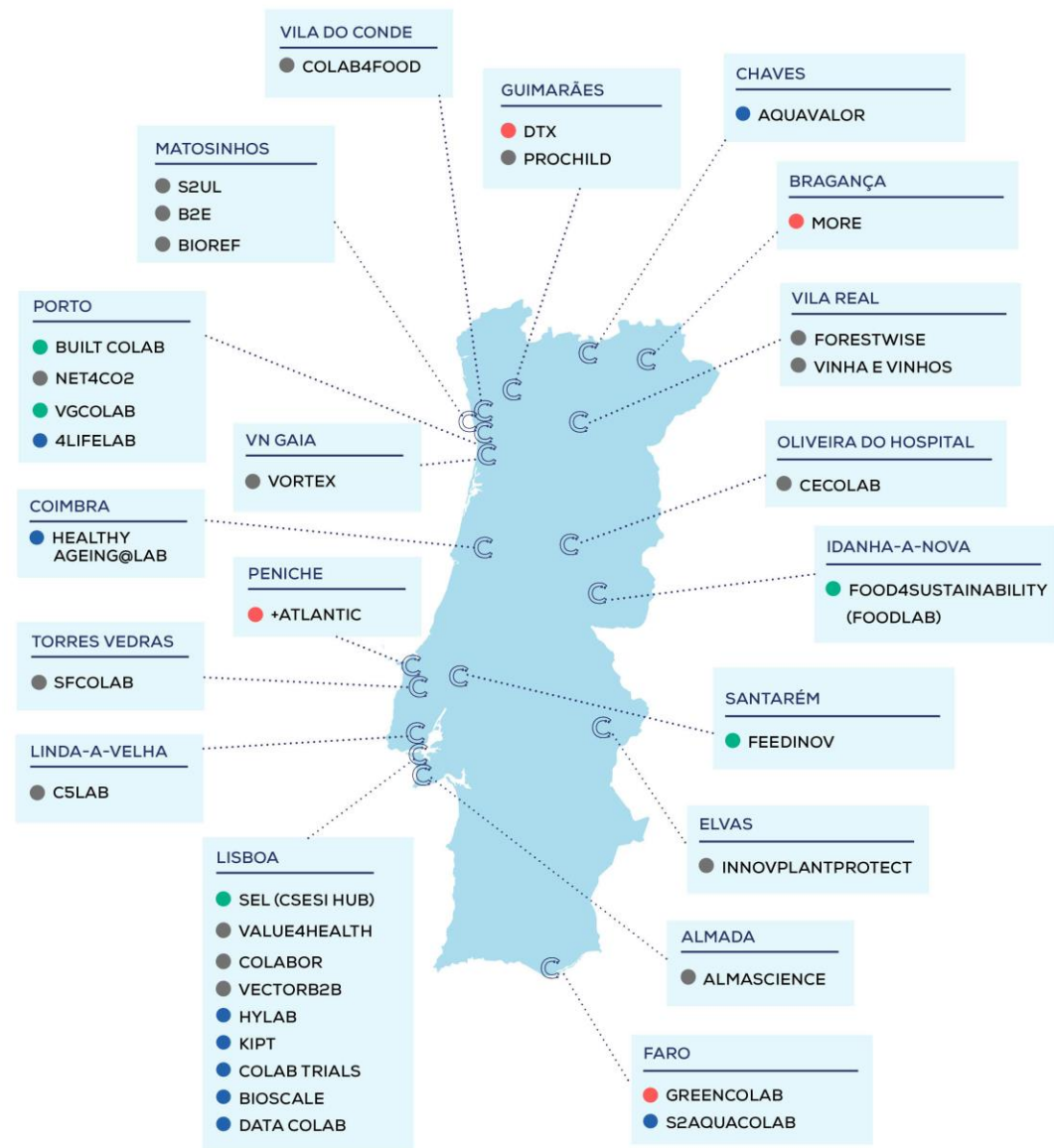
Digital
Transformation
CoLab

A large, abstract graphic composed of several thick, glowing lines in shades of blue and purple. On the left, a large chevron shape points towards the center. In the middle, a vertical bar is positioned above the text. On the right, a large 'X' shape is formed by two intersecting diagonal lines. At the bottom, a horizontal bar is positioned below the text. The background is a dark blue gradient with subtle light effects.

Experiencing
the Future

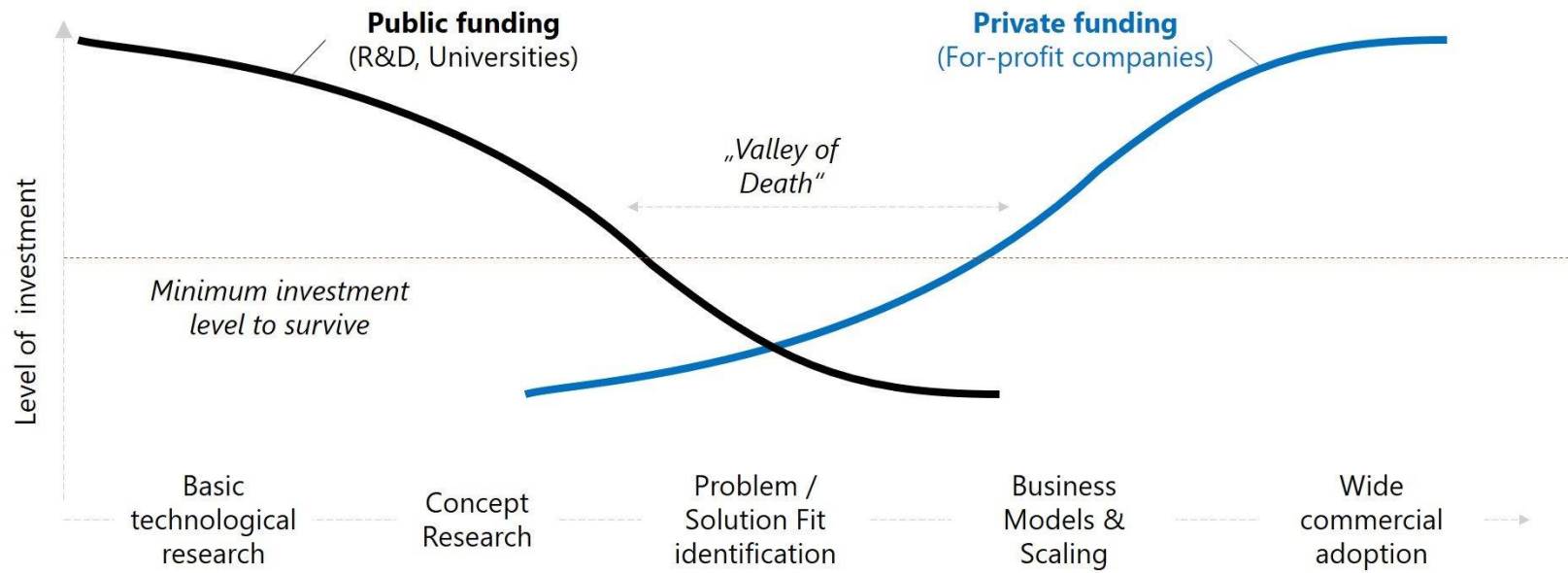
dtx-colab.pt

Rede de Laboratórios Colaborativos



● CoLAB reconhecidos em 2017 (1ª fase) ● CoLAB reconhecidos em 2019 (3ª fase)
● CoLAB reconhecidos em 2018 (2ª fase) ● CoLAB reconhecidos em 2021 (4ª fase)

Innovation Valley of Death



Associates

accenture

 **BOSCH**
Tecnologia para a vida

 **Cachapuz**
The best weigh 

CEIIA


celoplás®

 Centro de Computação Gráfica

cegid Primavera

 **SOCIETÀ COOPERATIVA
BILANCIAL**
Strumenti e Tecnologie per Pesare




INL
INTELLIGENT INJECTION
MAMOTECHNOLOGY
FOR MATERIALS

SIBS®

AERNnova

NOS

 **piep**
Inovação
na engenharia
de polímeros

dstgroup
building culture

 **GRUPPO
Simoldes**
Plastic Division

TMG
AUTOMOTIVE

 **UNIVERSIDADE
CATOLICA
PORTUGUESA**

 **UNIVERSIDADE
DE ÉVORA**


**Universidade do
Minho**

 **mobileum**
Action driven by intelligence



Software and Information Systems			Hardware and Sensors	Advanced Materials	Smart Manufacturing	Human Factors
Technological Innovation & Transfer Group Data Science and Machine Learning	Technological Innovation & Transfer Group Data and Application Engineering	Technological Innovation & Transfer Group Computer Graphics and Vision	Technological Innovation & Transfer Group Embedded and Edge Computing	Technological Innovation & Transfer Group Functional and Sensitive Materials	Technological Innovation & Transfer Group Process and Equipment Simulation	Technological Innovation & Transfer Group Ergonomics and Engineering Psychology
Future & Emergent Technologies Group						

Data Science and Machine Learning Group

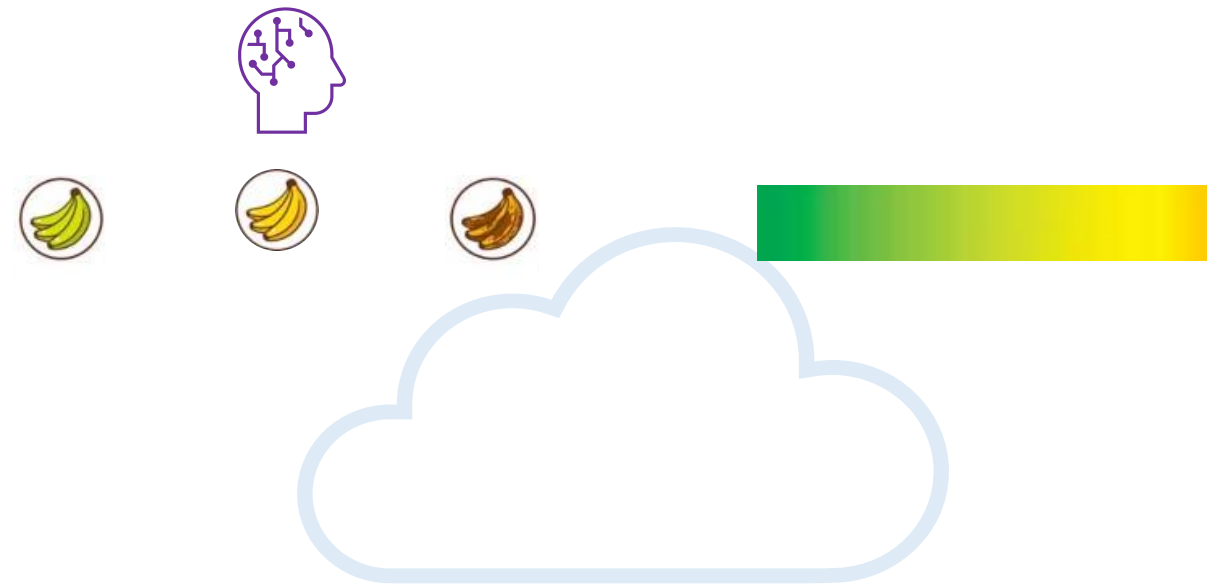


- 12 researchers
- 6 with PhD
- 5 with physics degree

Projects

The background features several abstract, glowing lines and dots in shades of blue and purple. The lines vary in thickness and orientation, some being straight and others slightly curved. The dots are small and circular, appearing as if they are the ends of the lines or separate points of light. The overall effect is a dynamic, futuristic, and somewhat ethereal composition.

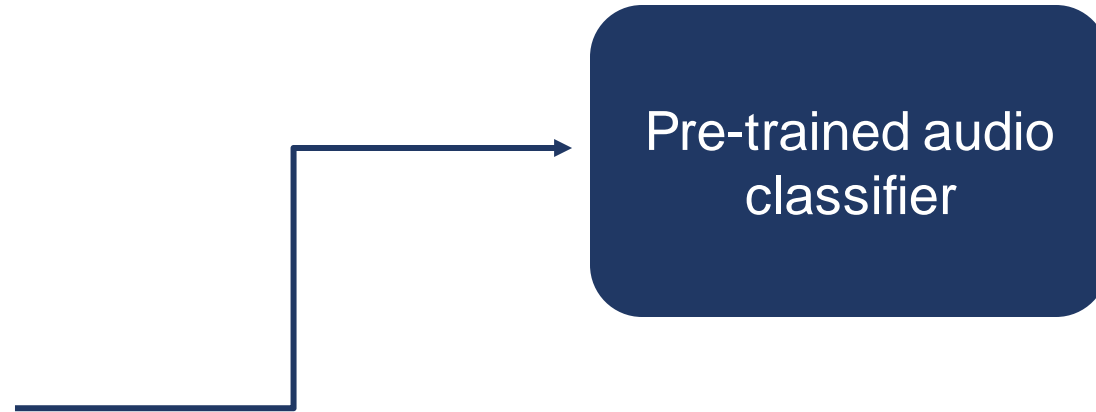
Automated optical inspection of perishable food



Highlight detection in audio of sports TV broadcasts



Highlight detection in audio of sports TV broadcasts

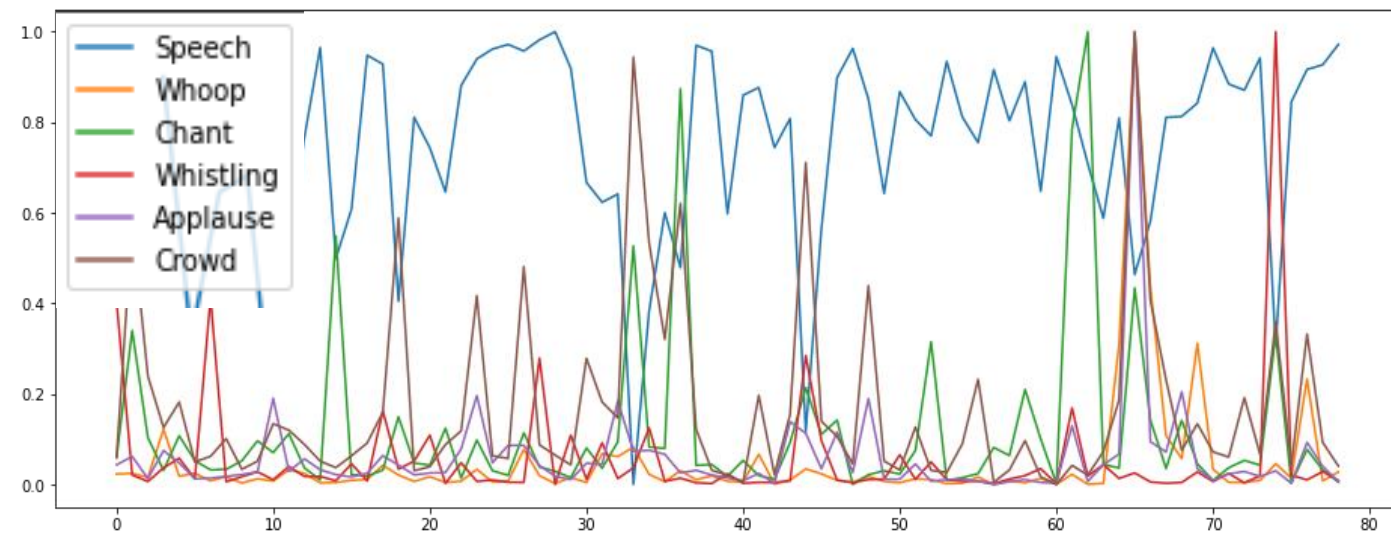


Pre-trained audio classifier

Highlight detection in audio of sports TV broadcasts



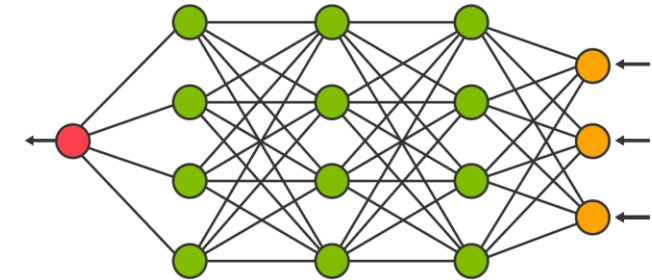
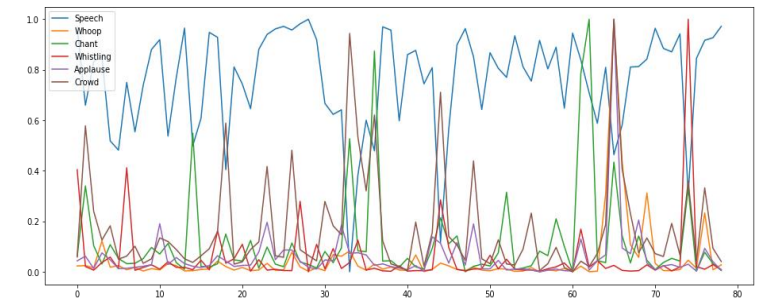
Pre-trained audio classifier



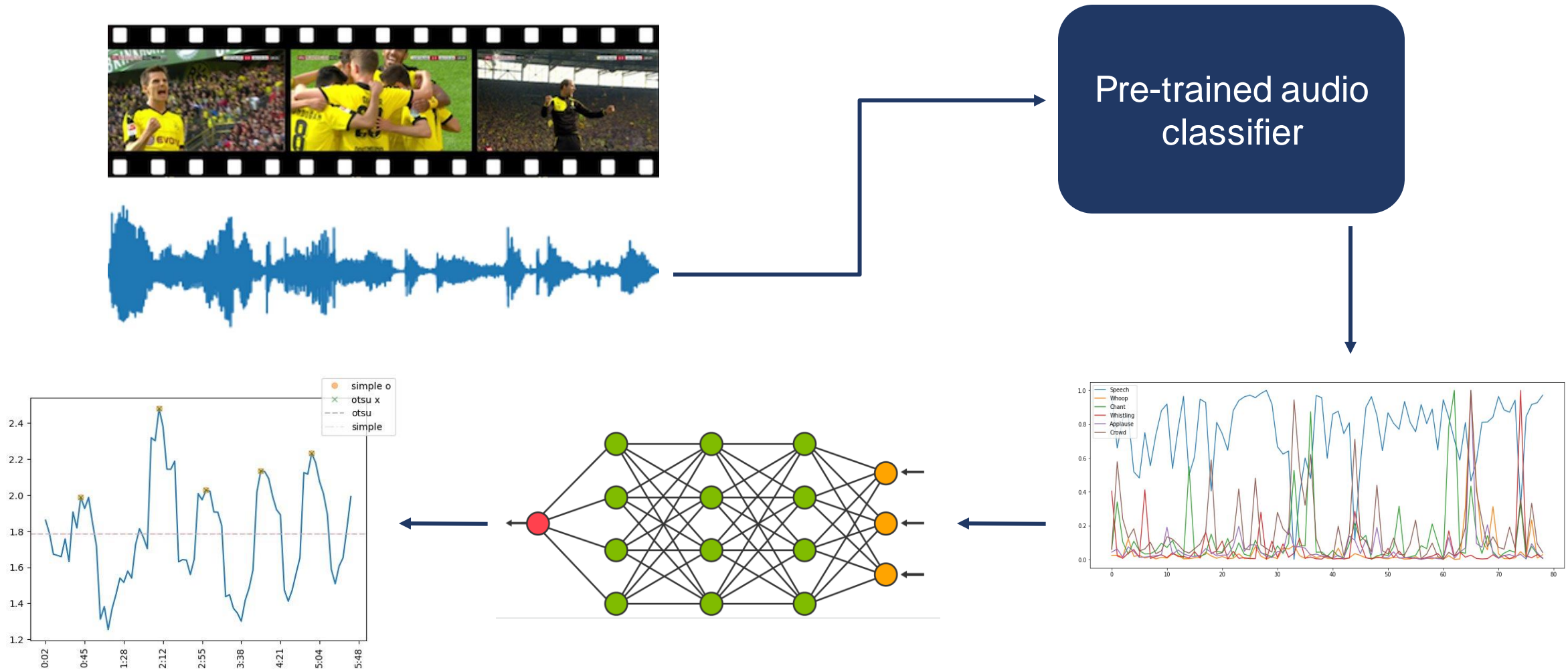
Highlight detection in audio of sports TV broadcasts



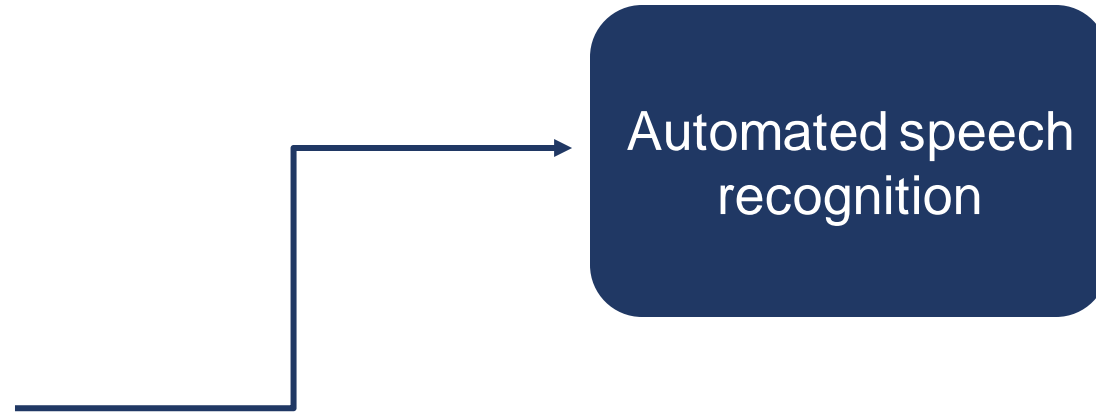
Pre-trained audio classifier



Highlight detection in audio of sports TV broadcasts



Highlight detection in audio of sports TV broadcasts



Automated speech recognition

Highlight detection in audio of sports TV broadcasts



Automated speech recognition

```
],  
"18": [  
  128.48,  
  131.3,  
  "Seems that an Ibrahimovic dropping off the front as he does a lot"  
],  
"19": [  
  132.18,  
  136.76,  
  "When Cavani plays that allows Cavani to spin into space at centre-forward"  
],  
"20": [  
  137.22,  
  139.22,  
  "Which can play the ball over the top for him"
```

Highlight detection in audio of sports TV broadcasts

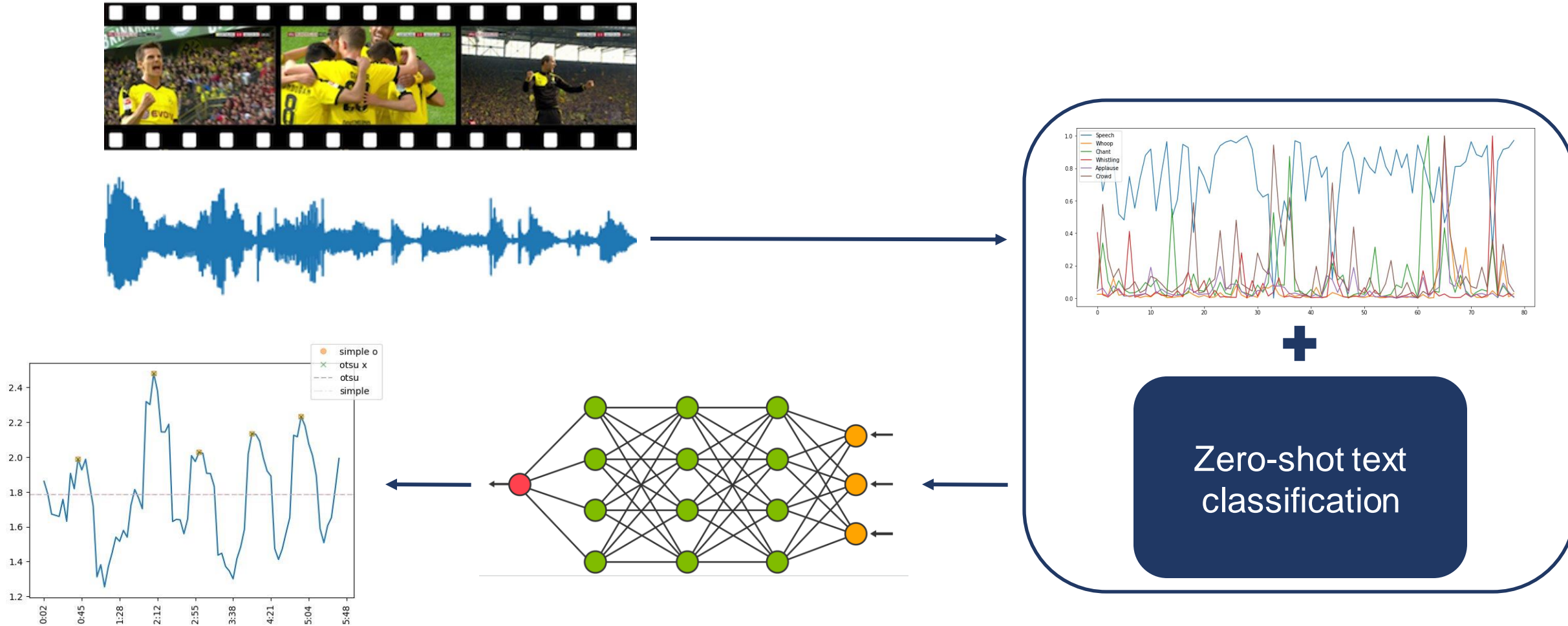


Automated speech recognition

Zero-shot text classification

```
],  
"18": [  
  128.48,  
  131.3,  
  "Seems that an Ibrahimovic dropping off the front as he does a lot"  
],  
"19": [  
  132.18,  
  136.76,  
  "When Cavani plays that allows Cavani to spin into space at centre-forward"  
],  
"20": [  
  137.22,  
  139.22,  
  "Which can play the ball over the top for him"
```

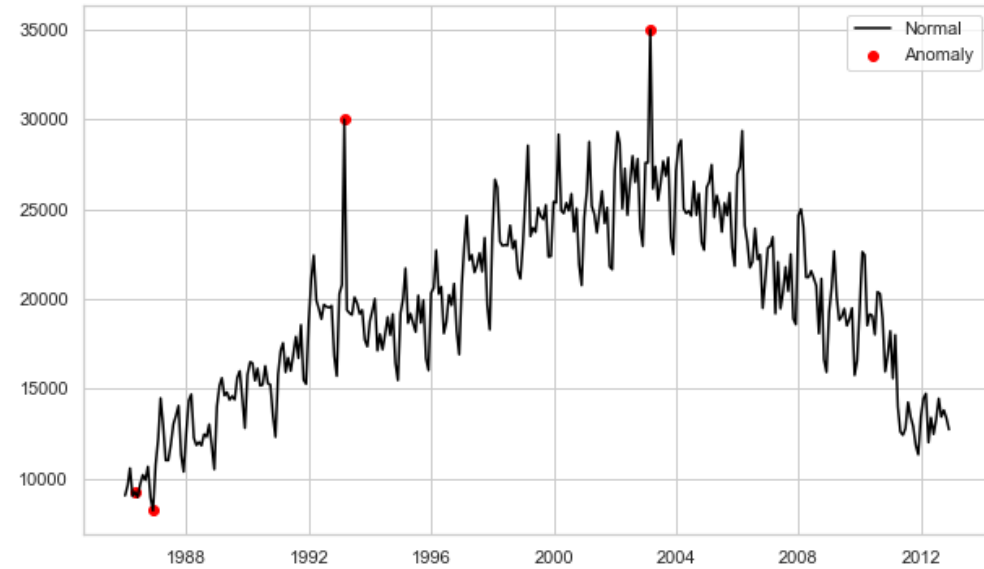
Highlight detection in audio of sports TV broadcasts



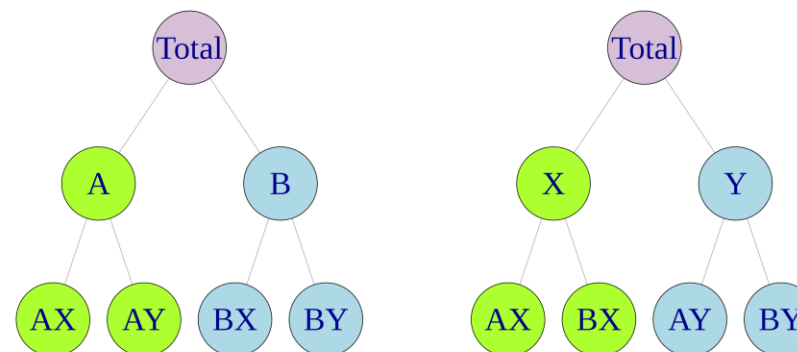
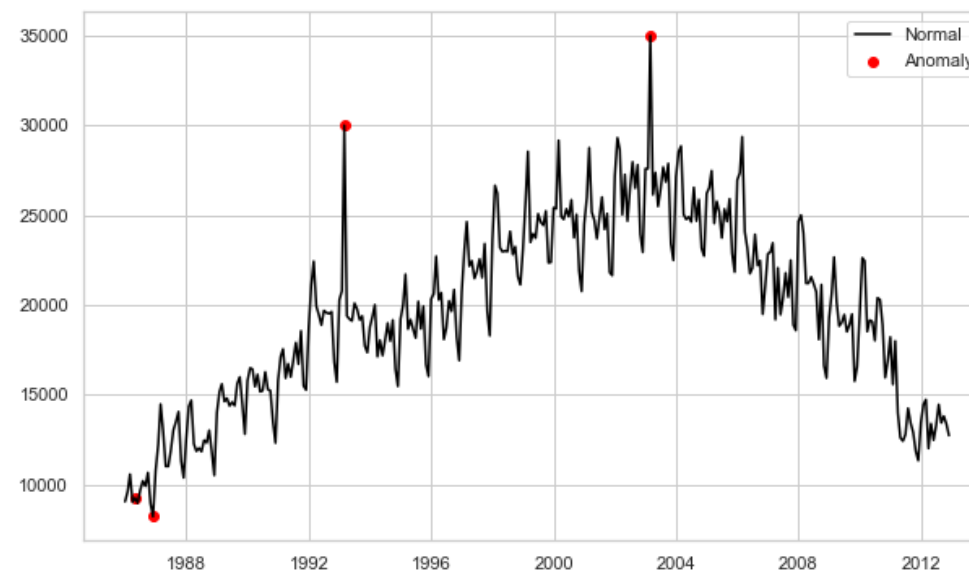
Anomaly detection in grouped time series



Anomaly detection in grouped time series



Anomaly detection in grouped time series



Anomaly detection in grouped time series

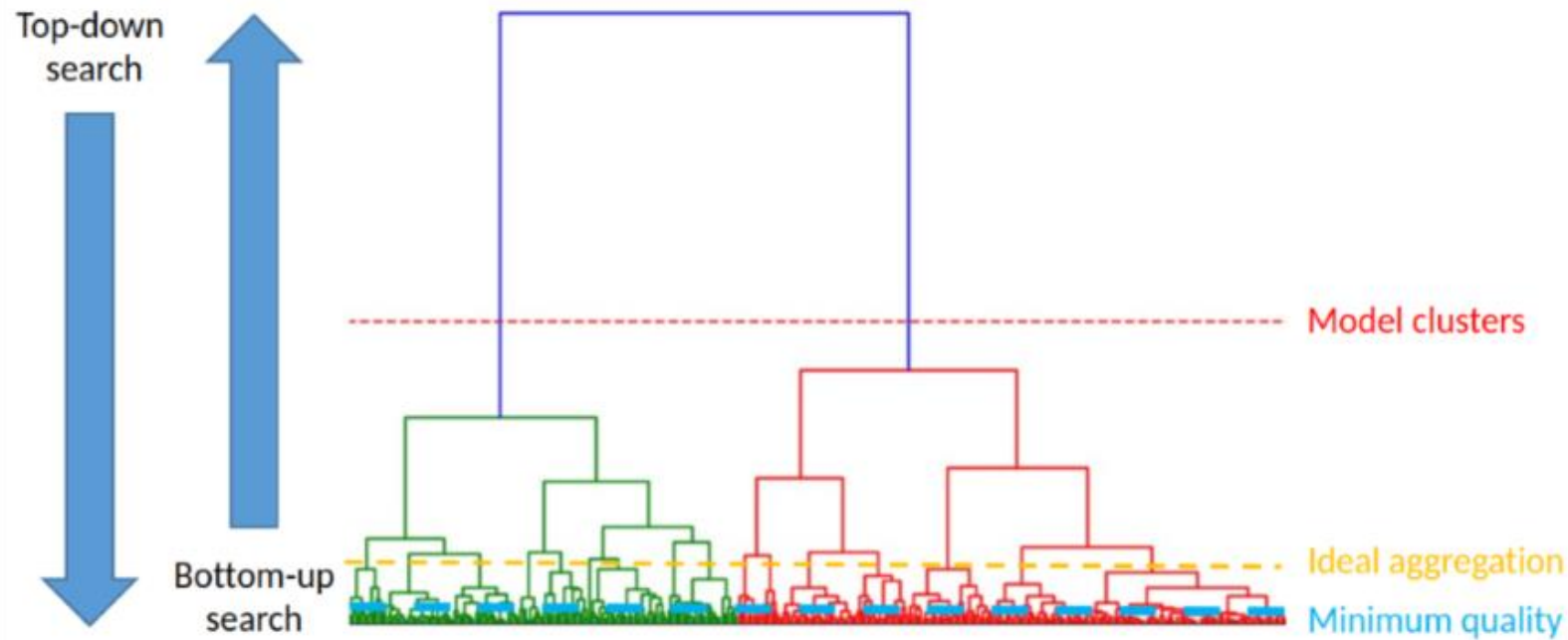
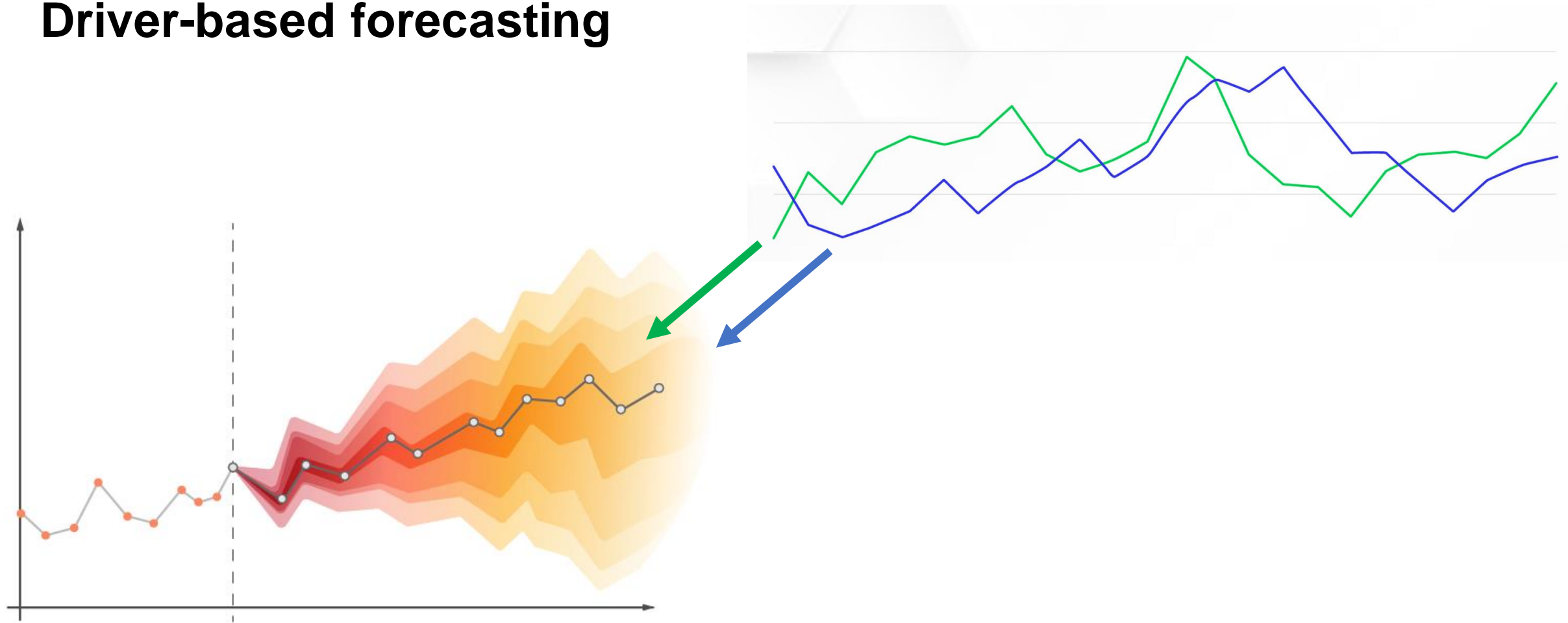
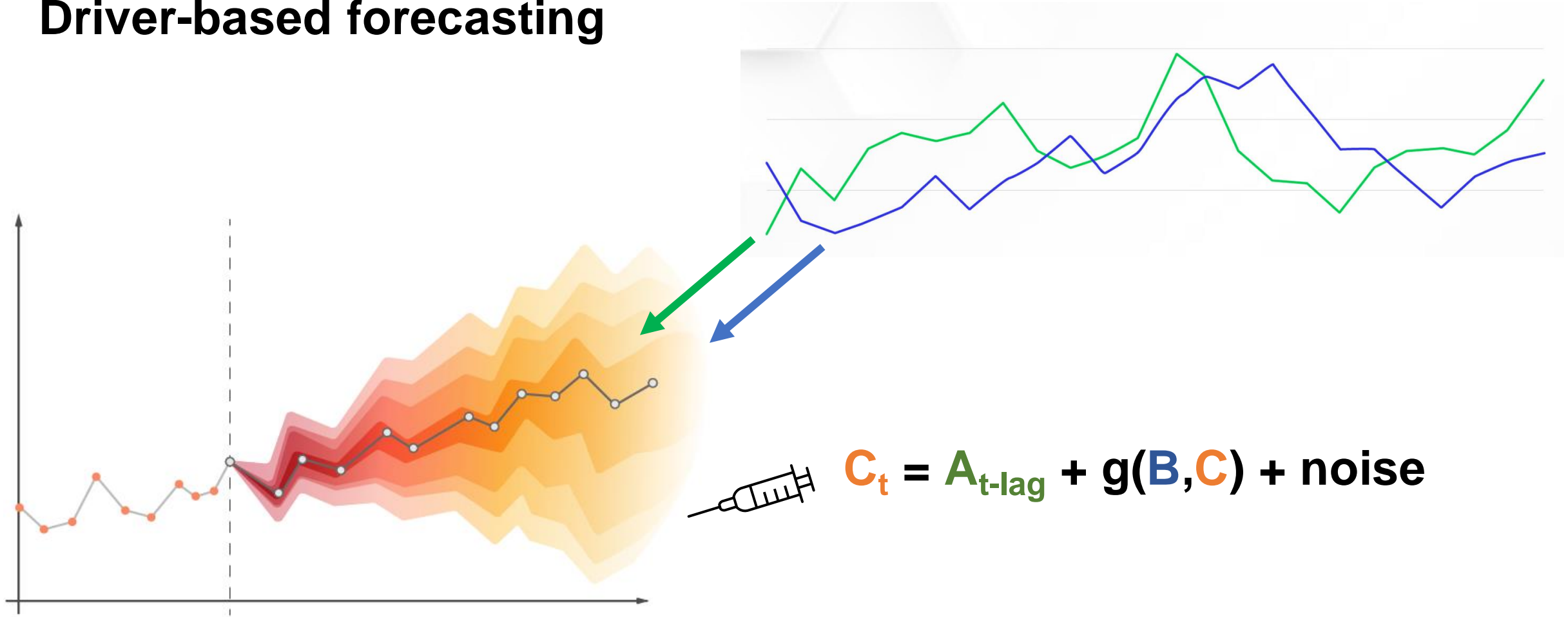


Figure 2 - Dendrogram representing the established hierarchical structure. The search for the level of aggregation can be bottom-up or top-down. The search is informed by the **quality criteria, expected anomaly detection performance and computational cost**, enabling to find the ideal model clusters and corresponding time-series (aggregated or not).

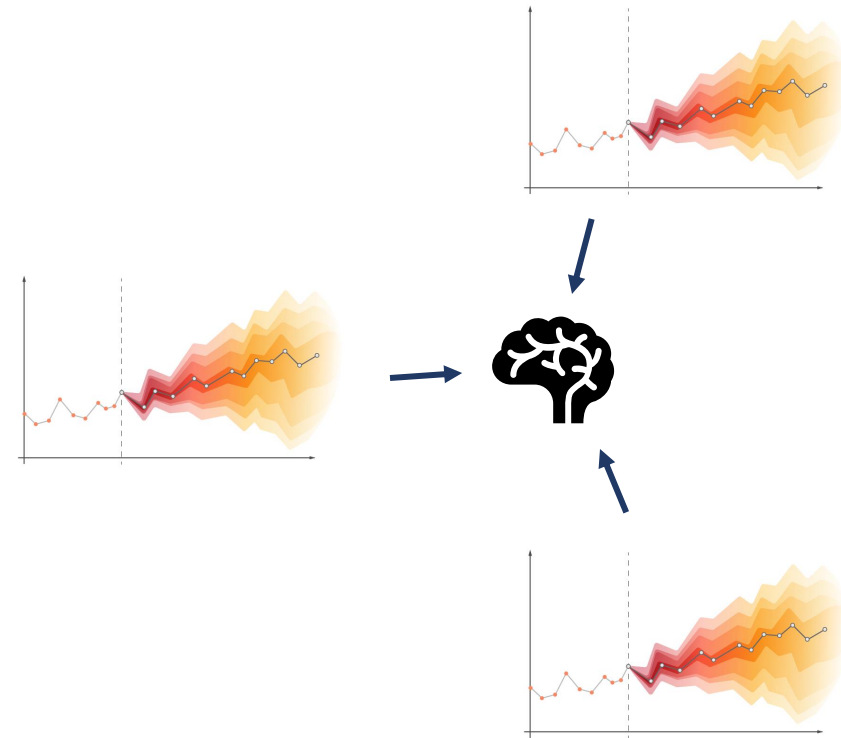
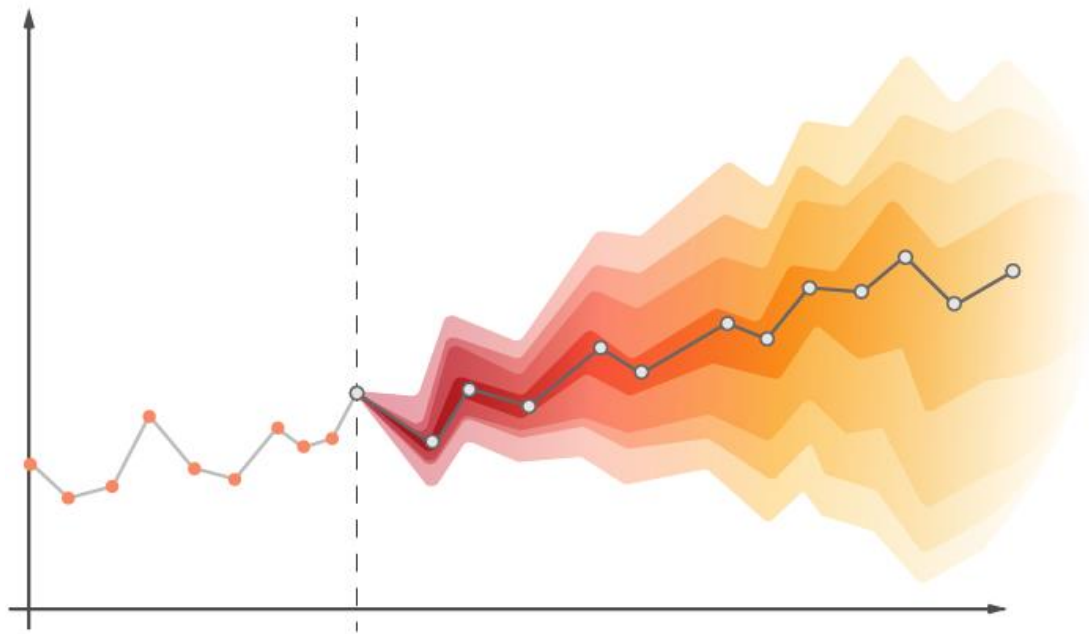
Driver-based forecasting



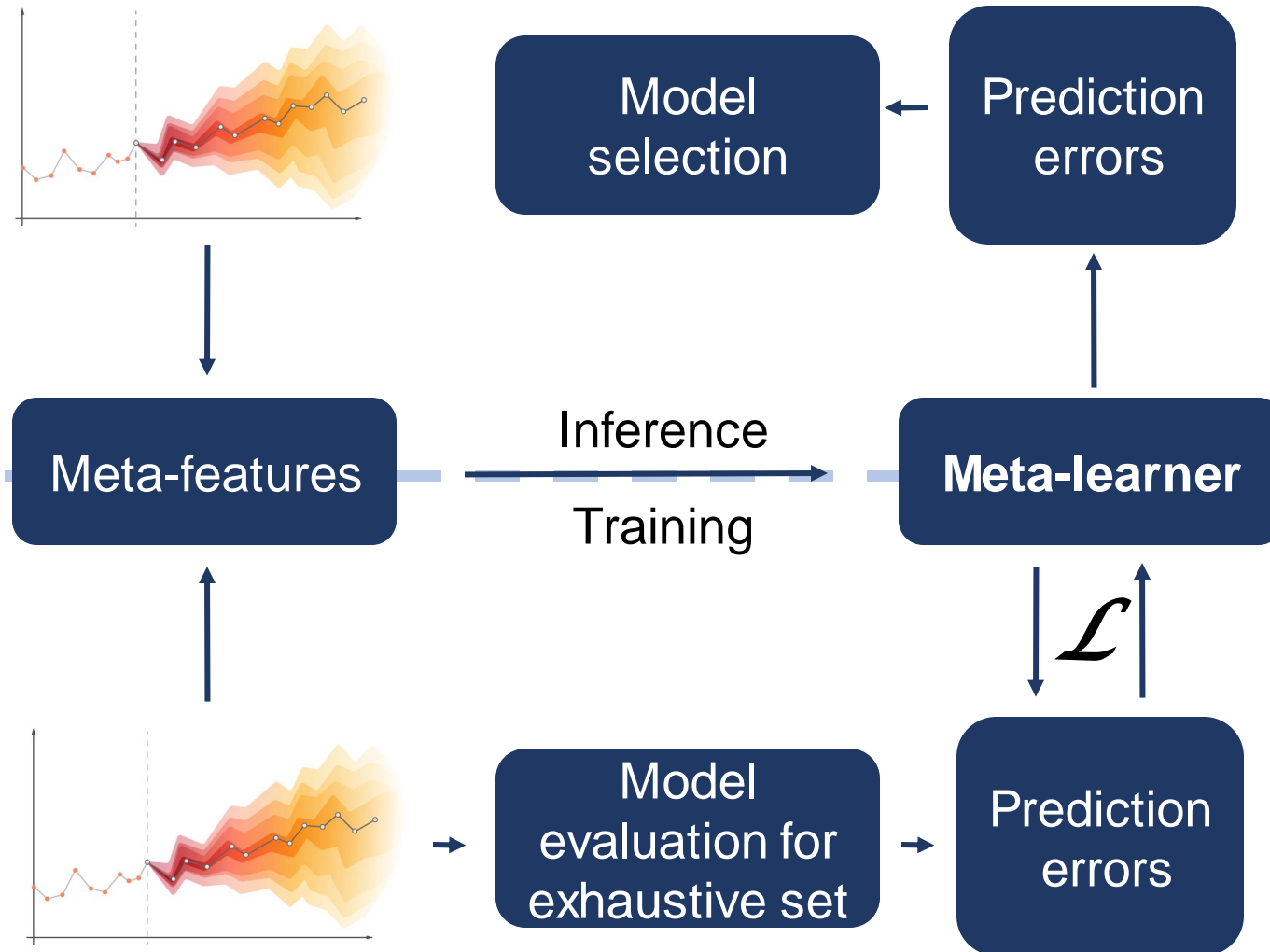
Driver-based forecasting



Driver-based forecasting



Driver-based forecasting

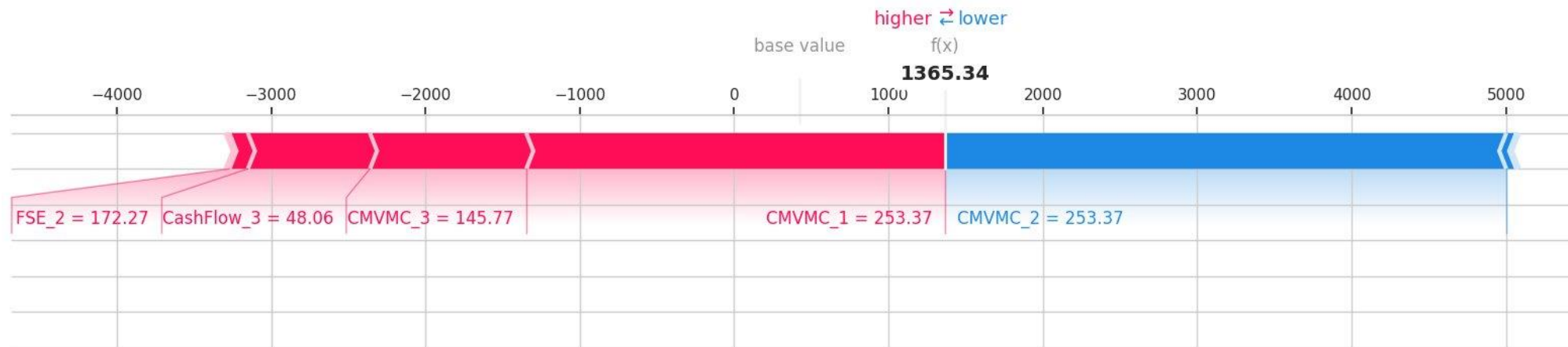


Driver-based forecasting

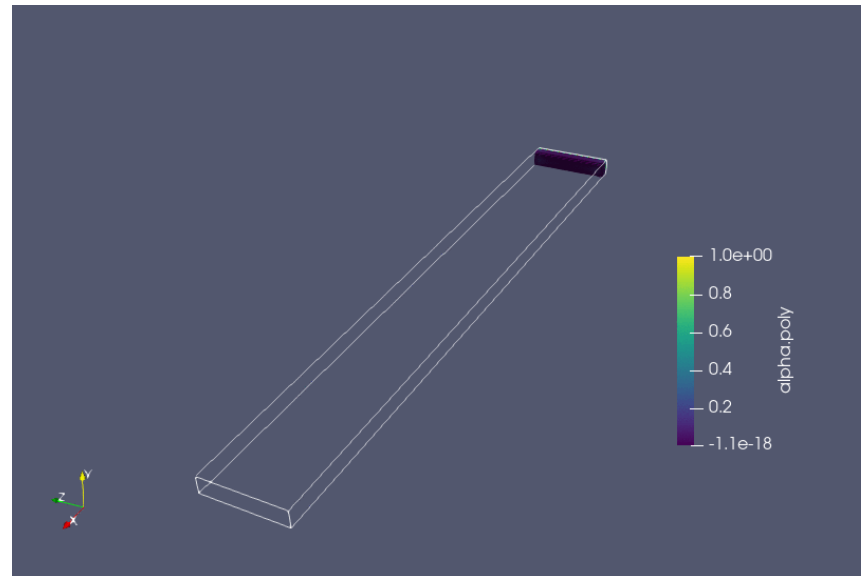
Feature Importance



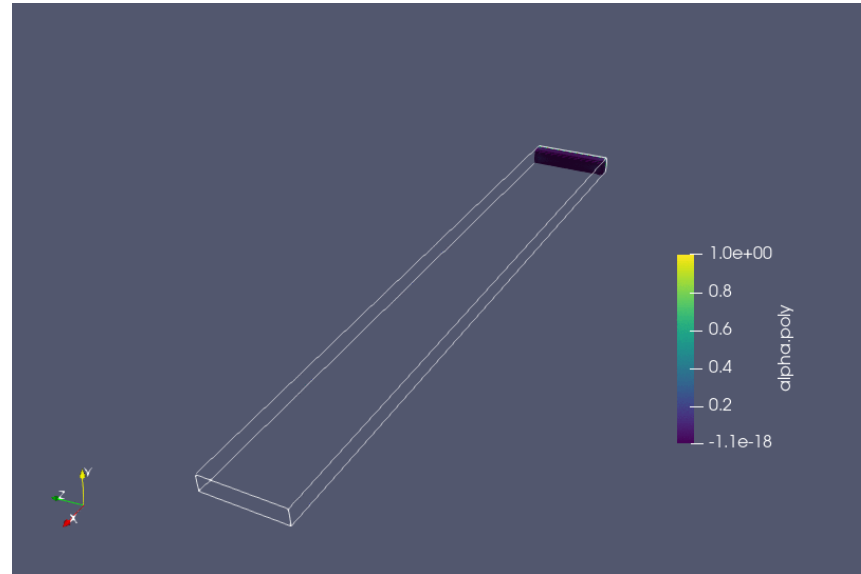
Shap Force Plot for Horizon: 2



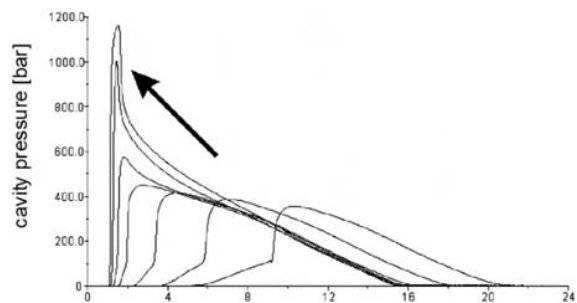
Industrial process control



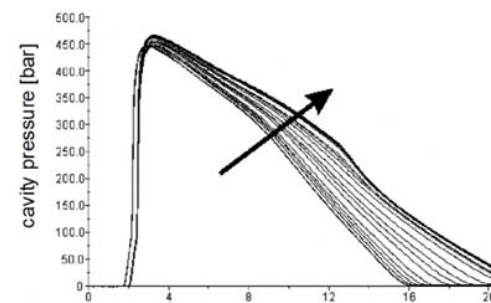
Industrial process control



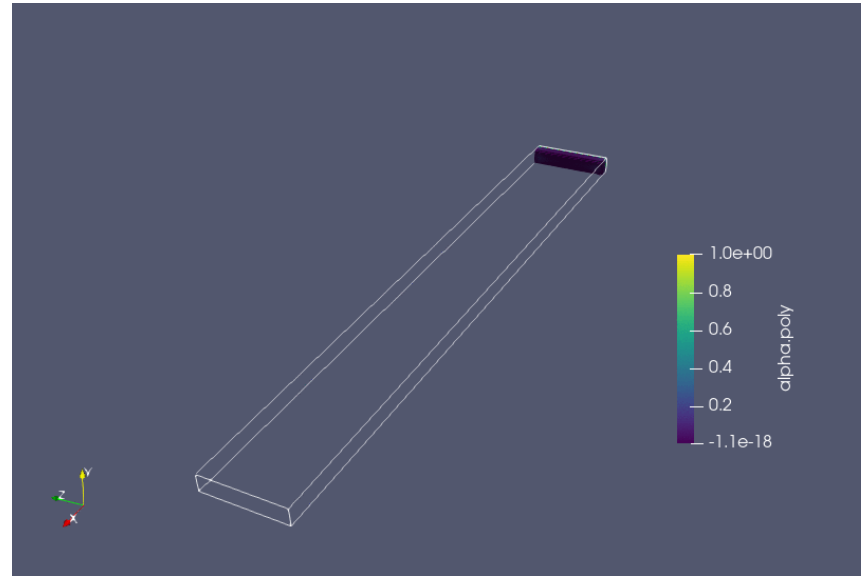
Injection Speed



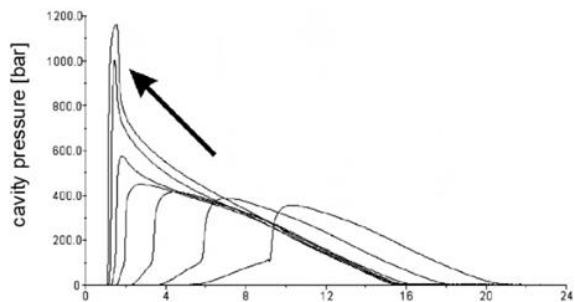
Mold Temperature



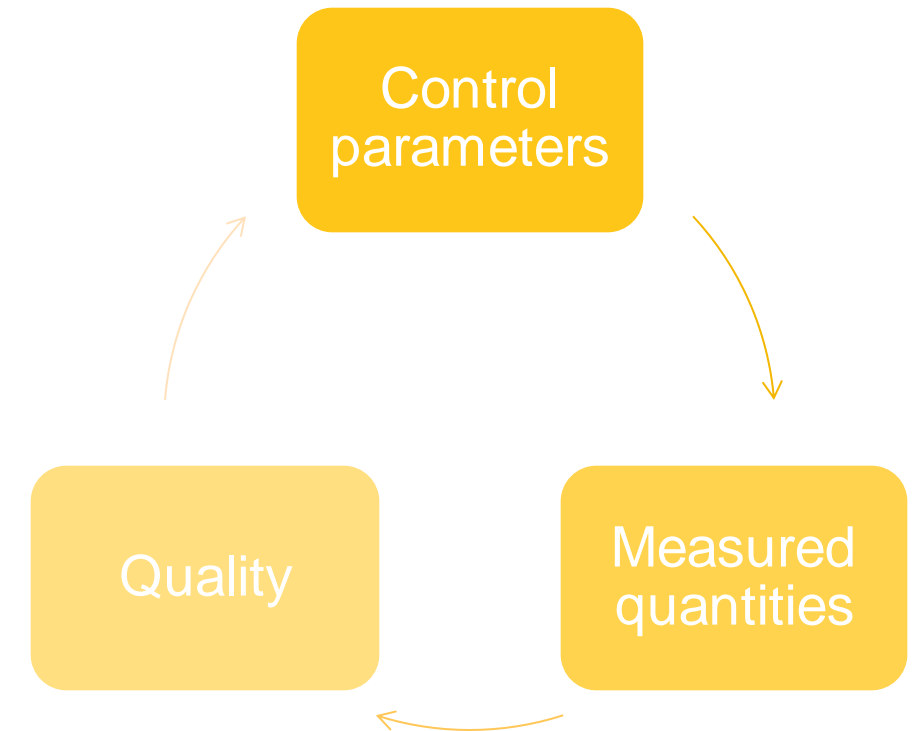
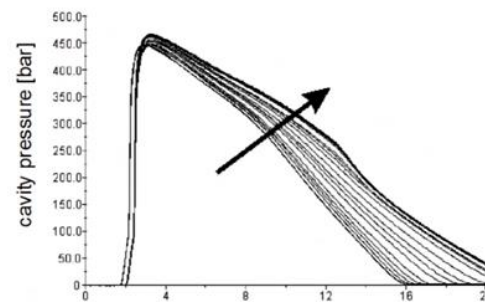
Industrial process control



Injection Speed



Mold Temperature



Final remarks

The background features several thick, diagonal lines in shades of blue and purple, creating a sense of movement and depth. There are also several small, solid-colored circles scattered across the scene, some appearing to be at the ends of the lines. The overall aesthetic is modern and minimalist.

Final remarks

- Interface organizations like DTx provide unique environment to grow as data scientist
 - Diverse interesting problems from industry
 - Exposure to network of clients/associates
 - Focus on economic value creation
 - Still with more room for experimentation than in companies
- How physics training can differentiate you in data science
 - Not too focused on particular tools
 - Intuition to look for what can be neglected and to make approximations
 - Confidence to come up with new approaches from first principles, or to adapt established methods

Final remarks

- Potential pitfalls:
 - Physicists have a talent for finding first-principle reasons to conclude that something is not worth trying
 - "The data is not good/enough" - some times it's true, other times it's a lazy response. Consider injecting domain knowledge or using transfer learning, data augmentation, dimensionality reduction, ...
 - Ignoring the domain knowledge. As a data scientist, you will not become a domain expert for every project, but being curious about the domain will only give you an advantage



emanuel.gouveia@dtx-colab.pt