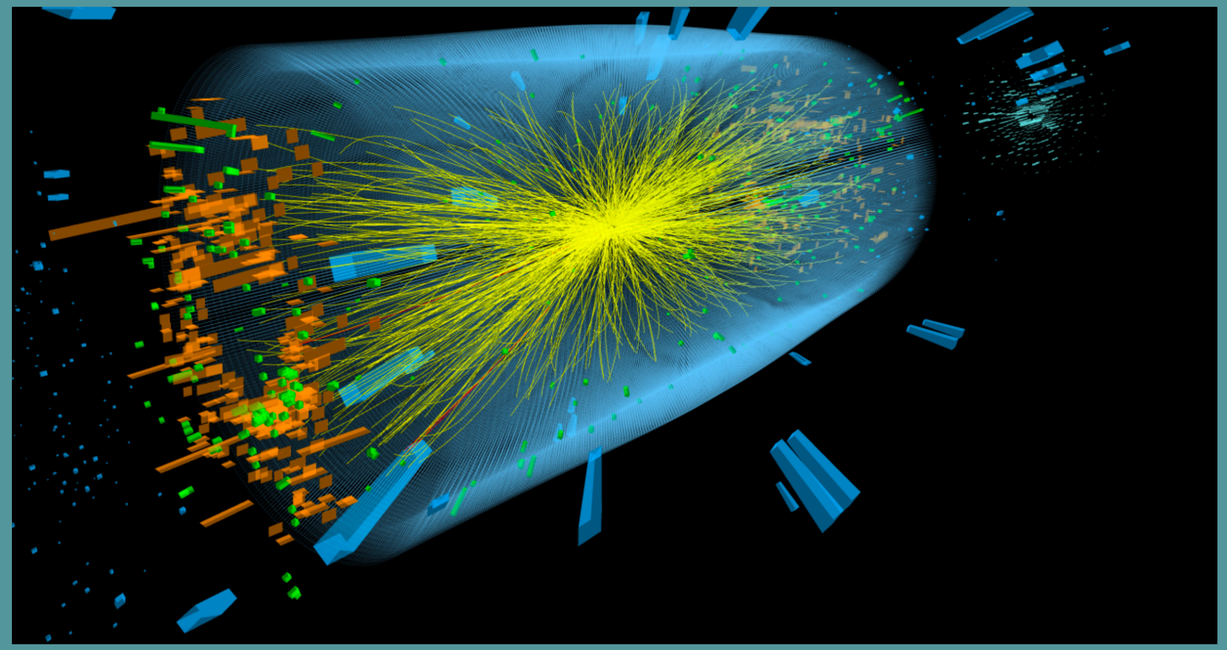
Portfolio Projects

Olympia Dartsi

Research at CERN

(European Organization for Nuclear Research)



To test our best theories about the fundamental structure of our universe, CERN has build a massive particle collider which generates an enormous amount of data:

~1 billion of collisions/second -> 1PB of collision data per second

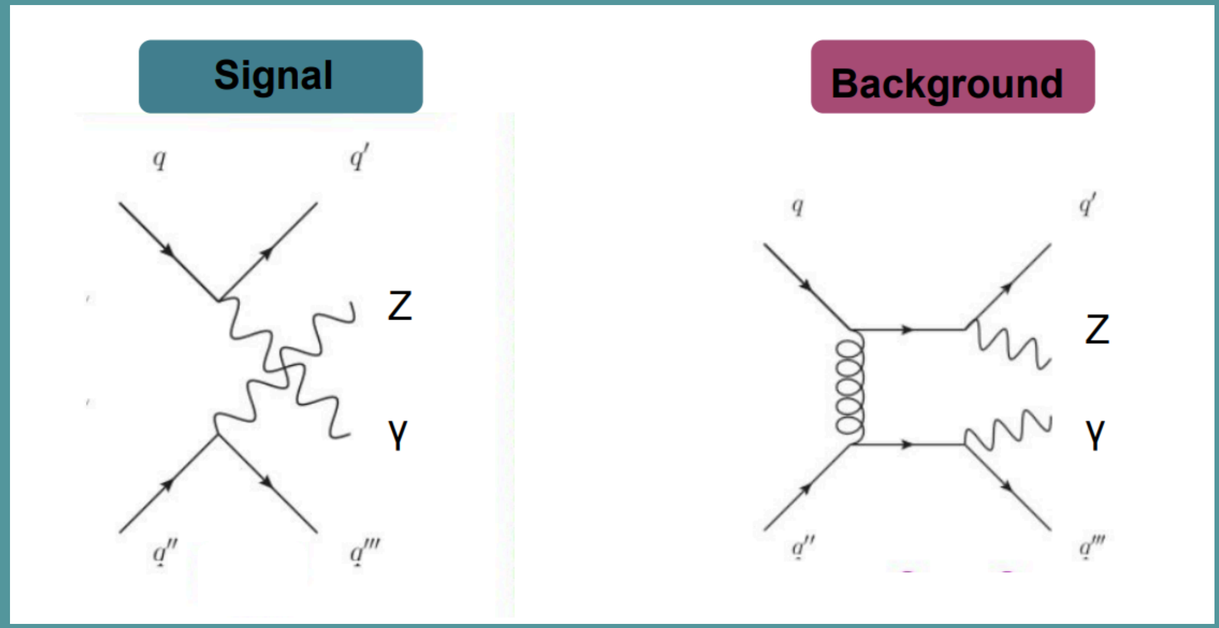
Hundreds of different processes happening all at once

Challenge for researchers: Extract the signal they are interested in by eliminating the background noise

Research at CERN

Context: The Standard Model predicts a mechanism for proton decay that has not yet been experimentally confirmed.

- **Problem 1**: there is another decay with same initial and final state that is much more likely to happen.
- **Problem 2**: the two processes interfere, generating additional noise
- **Problem 3**: detector limitations add bias to the results

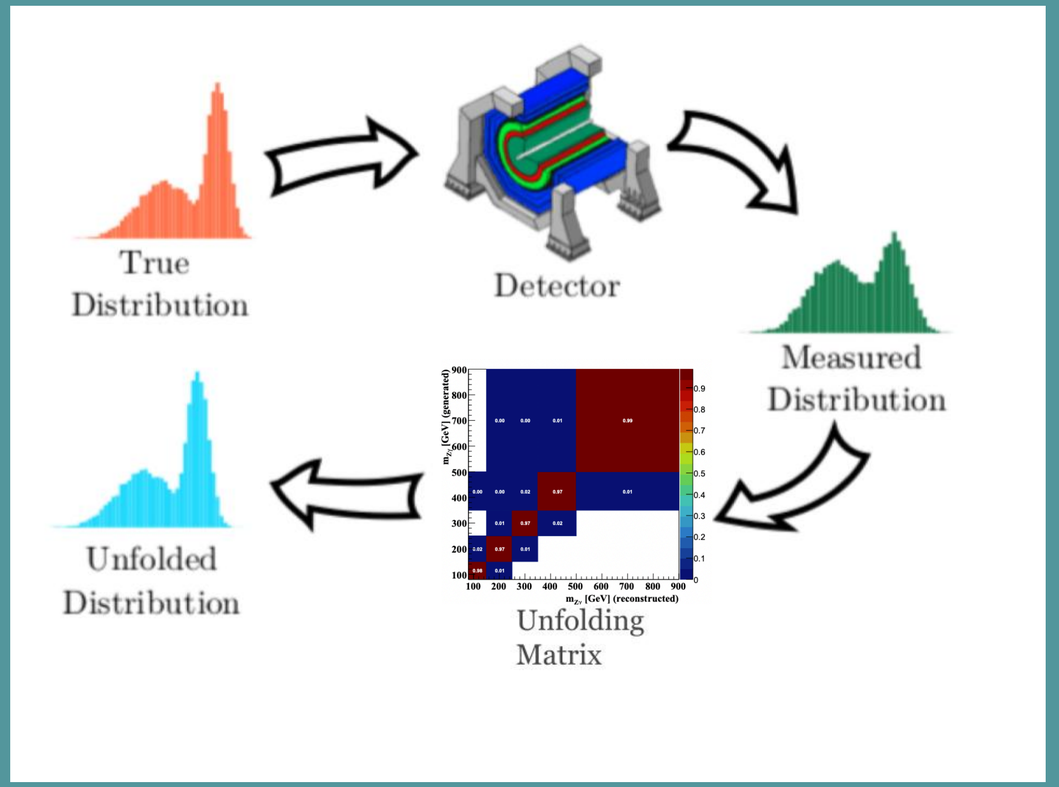


Challenge: How do we figure out which events in our data correspond to the process we are interested in?

Research at CERN

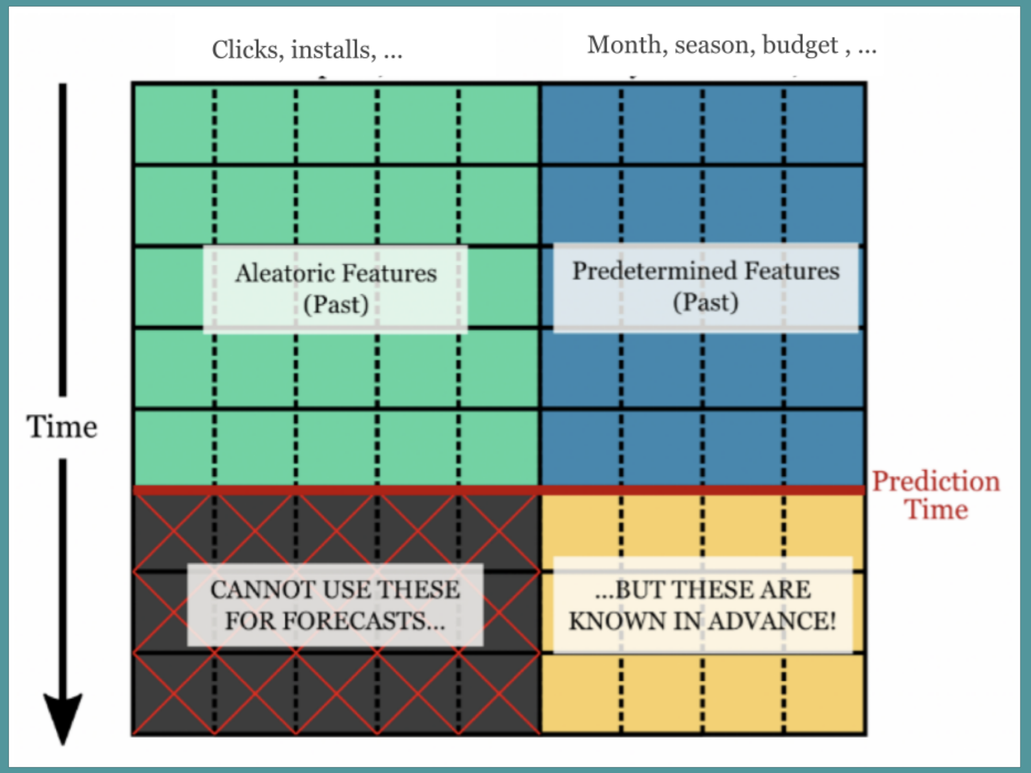
How do we address these problems?

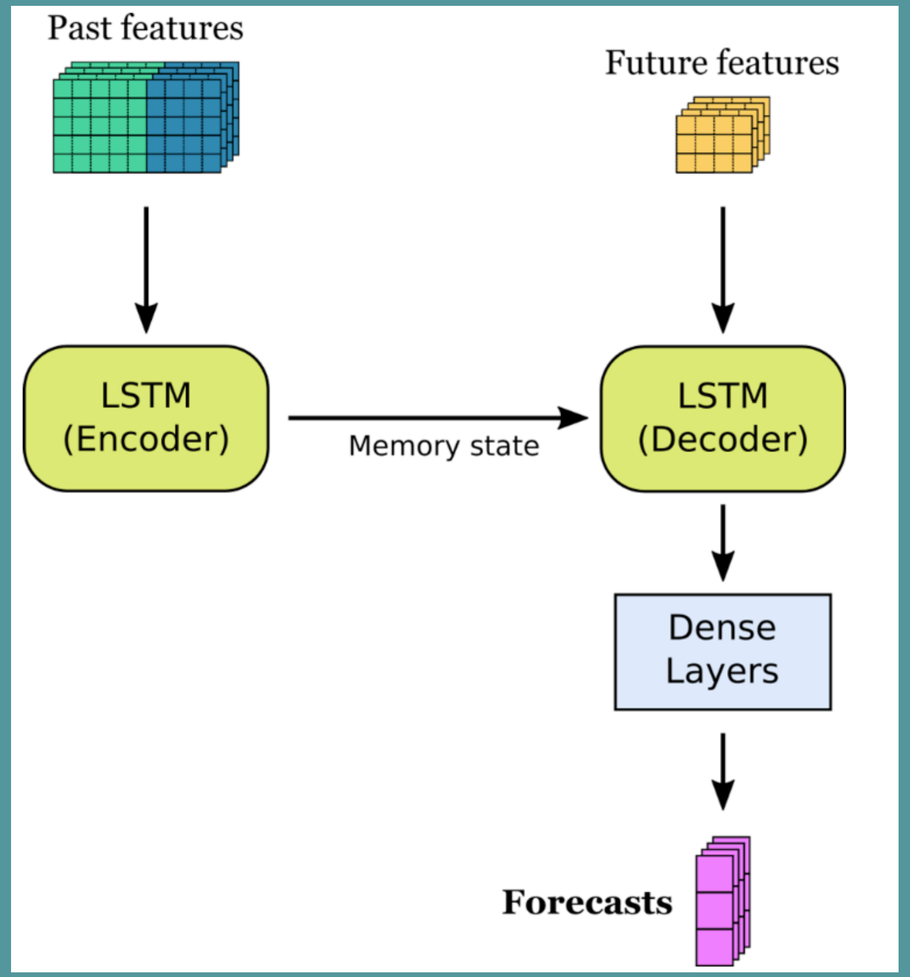
- **Solution for 1**: Characteristics of final states differ (example: mass of electrons).
  -> Build a **boosted decision tree** to separate the contributions
- **Solution for 2**: Use Monte Carlo simulation to estimate magnitude of interference effects
- **Solution to 3**: Calibrate raw data to account for detector imperfections using an “unfolding” method based on Bayes’ theorem



Time series forecasting using Neural Network

- Purpose: to predict future revenues
- Based on an Encoder-Decoder architecture
- Encode categorical variables (year, month, day, country, holiday, etc.) with a One-Hot encoder.





Computer Vision projects

- Unsupervised anomaly detection using auto encoders for MNIST digit dataset
- Training data: no anomalies
- Test data: both normal and irregular digits
- Assumptions:
  - After encoding and decoding, the training images will look almost identical
  - For outliers in test data, original and reconstructed images will differ