POLITECNICO
MILANO 1863

Towards Time-Evolving Analytics: Online Learning from Time-dependent Evolving Data Streams

Giacomo Ziffer

PhD Student, DEIB – Politecnico di Milano

05/12/2022

Agenda

- Adaptive Machine Learning for Data Streams
- Temporal Dependence in Data Streams
- Time-Evolving Analytics
- Contributions (so far)
- Ongoing and Future Works

Adaptive Machine Learning

MIT Technology Review

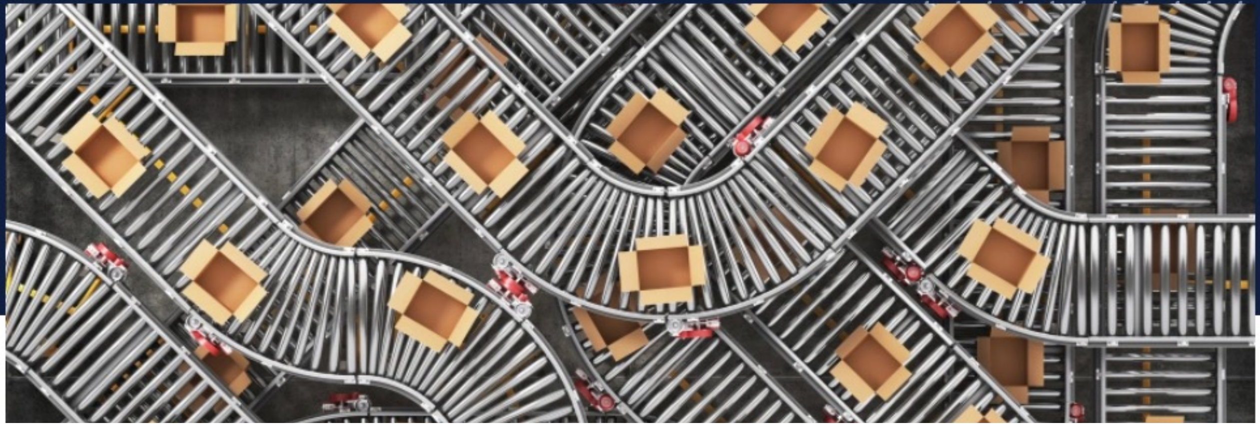
Featured Topics Newsletters Events Podcasts My account Subscribe

ARTIFICIAL INTELLIGENCE

Our weird behavior during the pandemic is messing with AI models

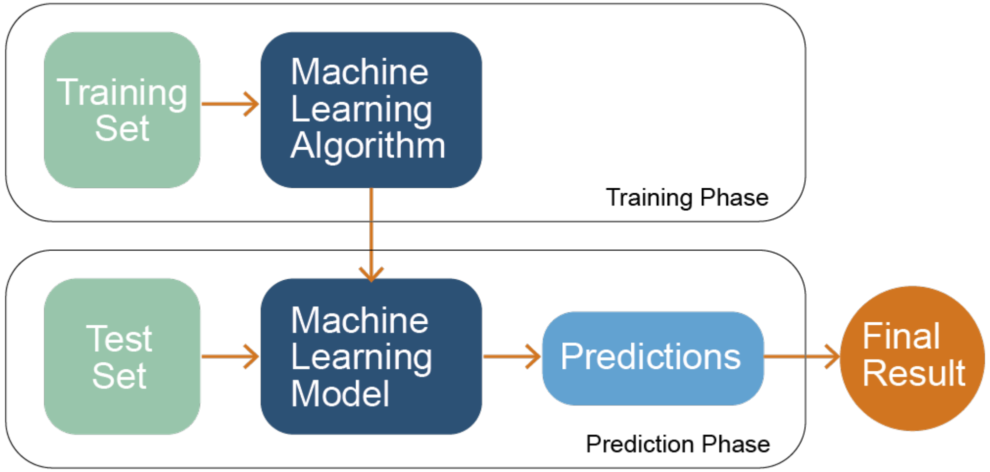
Machine-learning models trained on normal behavior are showing cracks — forcing humans to step in to set them straight.

By Will Douglas Heaven May 11, 2020

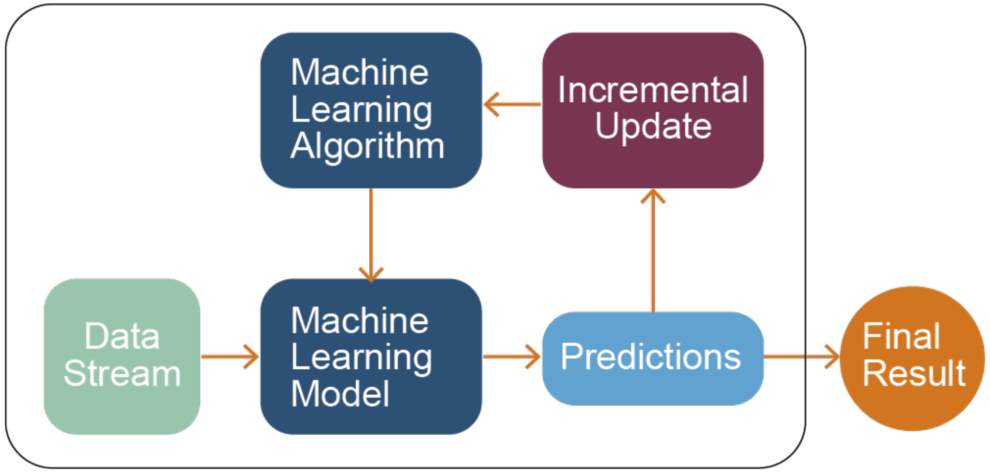


Offline Learning vs Online Learning

Offline Learning



Online Learning

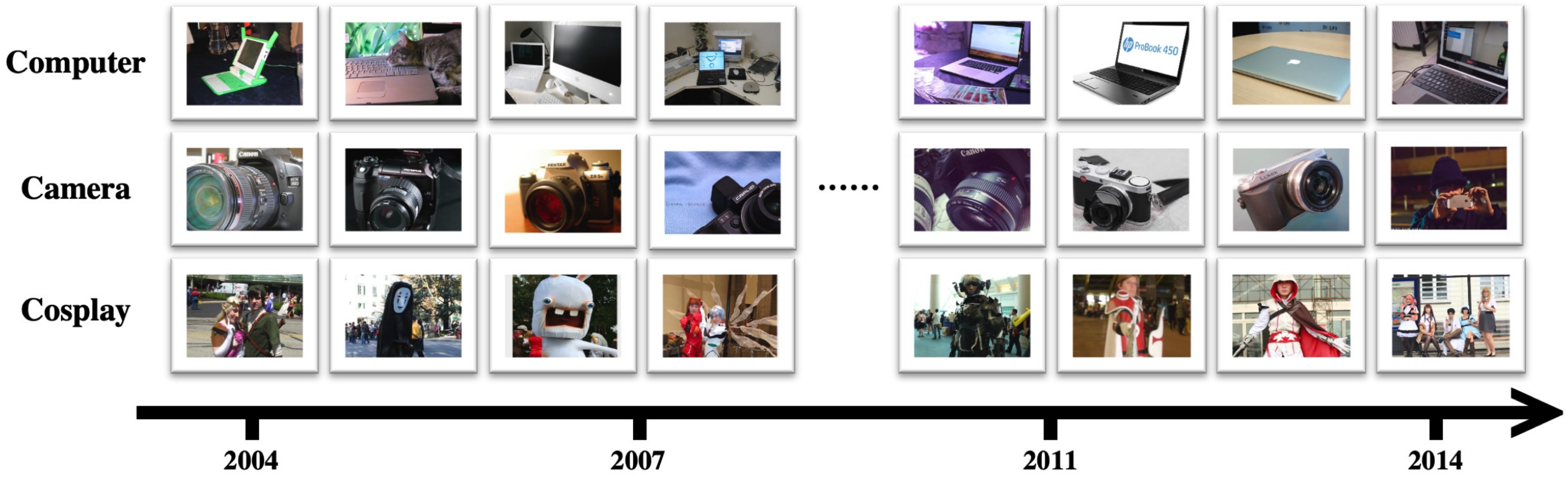


i.i.d. assumption in data stream

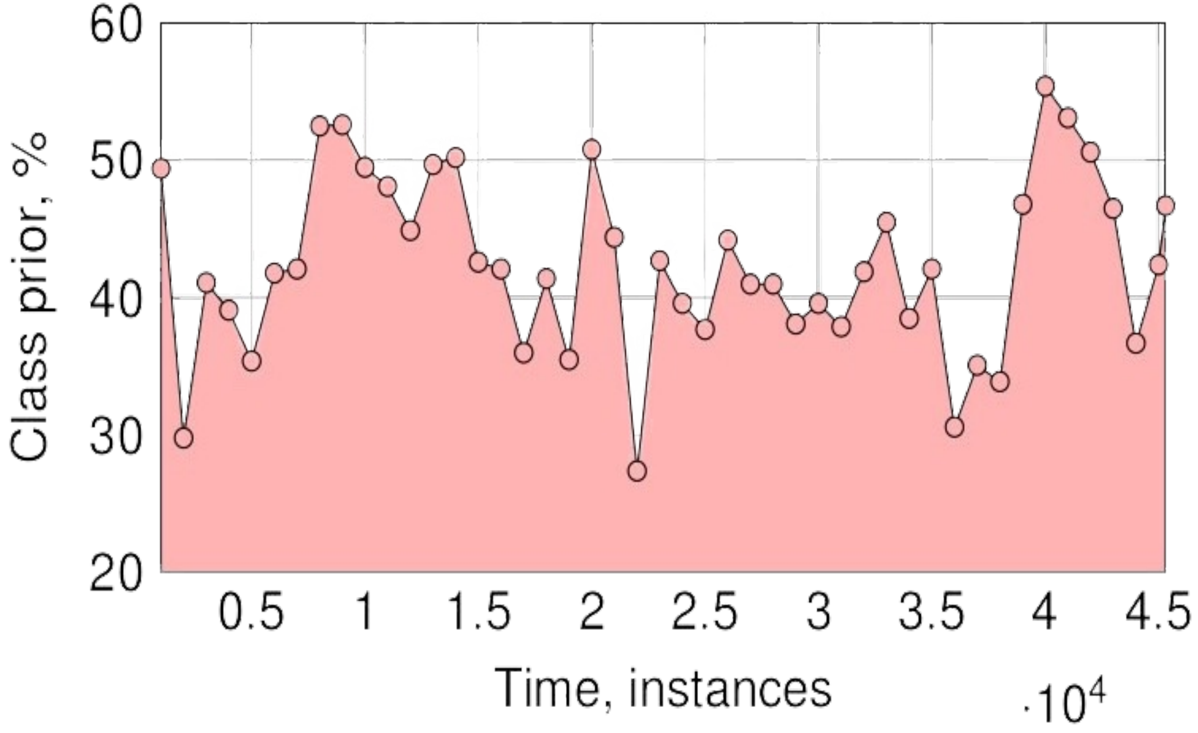
Typical assumption (in data streams):

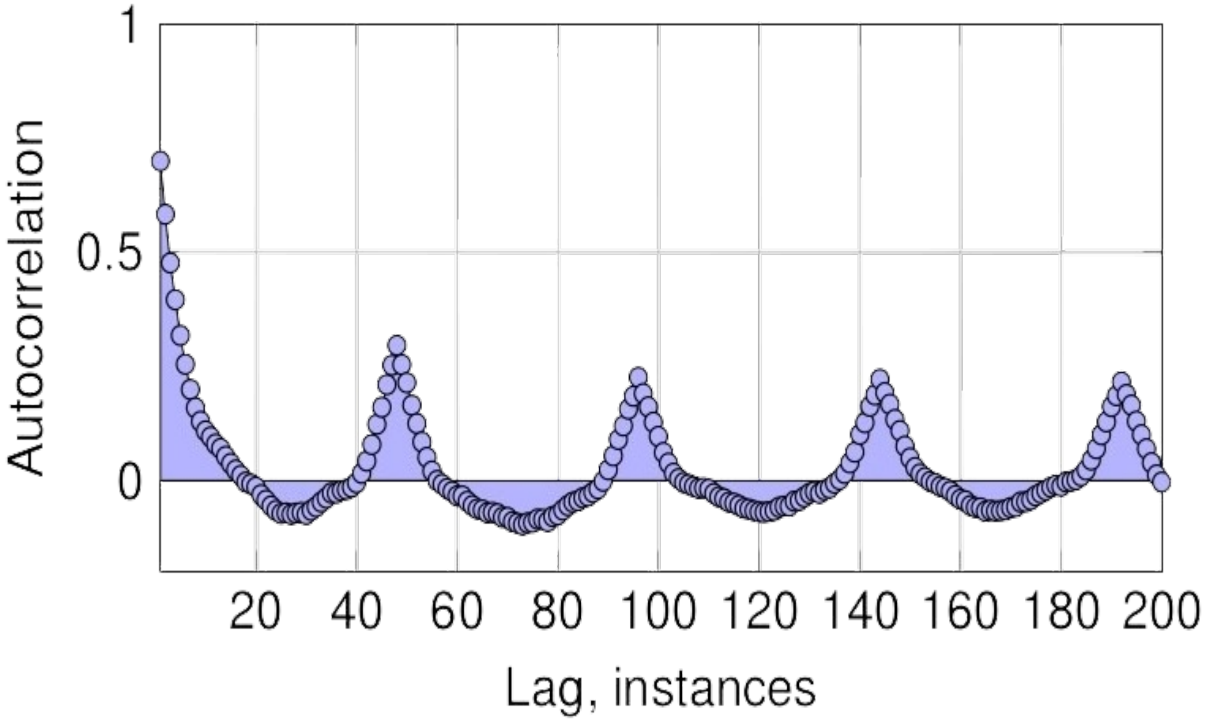
$$(x_t, y_t) \sim P_t \quad \text{is iid wrt } P_t$$

i.e., no temporal dependence and same distribution **per concept**. So, drift is something to detect and 'deal with' , by rebuilding/adapting models.



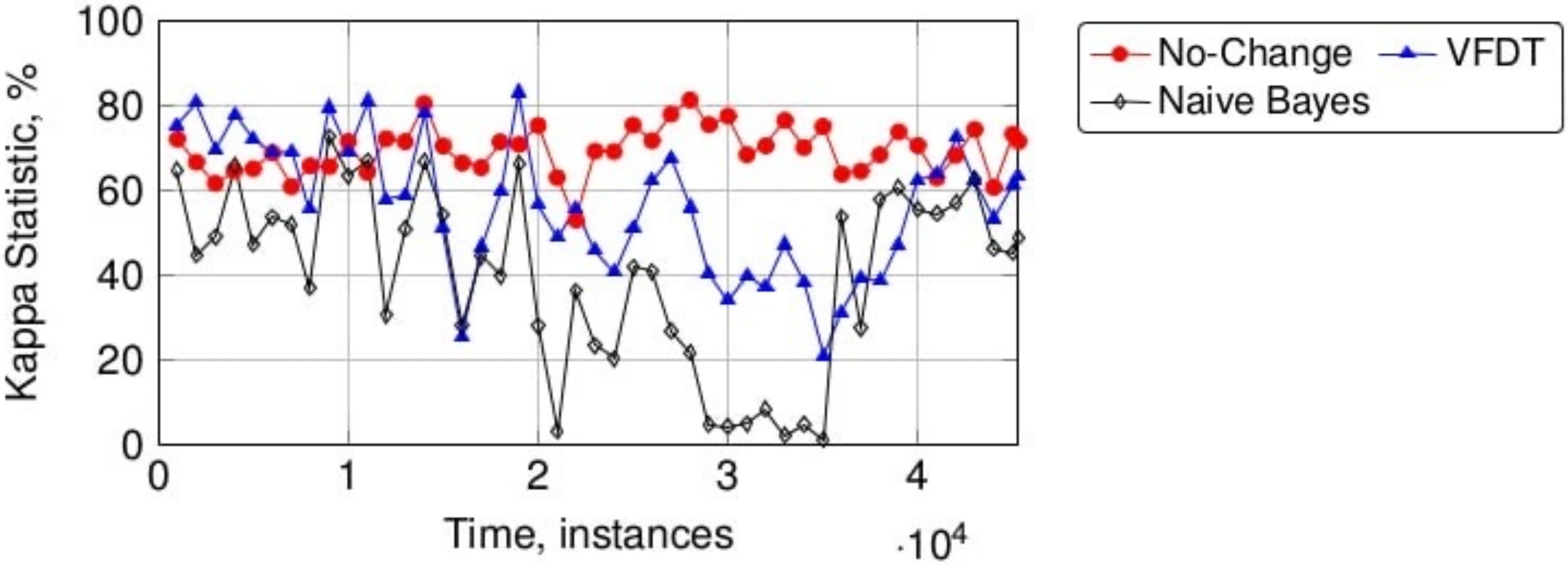
i.i.d. assumption vs. the Electricity data stream





Bifet, Albert, et al. **Pitfalls in benchmarking data stream classification and how to avoid them.** ECML PKDD. Springer, Berlin, Heidelberg, 2013.

Electricity data stream and Kappa Statistic



Bifet, Albert, et al. **Pitfalls in benchmarking data stream classification and how to avoid them.** ECML PKDD. Springer, Berlin, Heidelberg, 2013.

Existing solutions

Comparison among Offline Machine Learning, Time Series Analysis, Incremental Learning, and Streaming Machine Learning

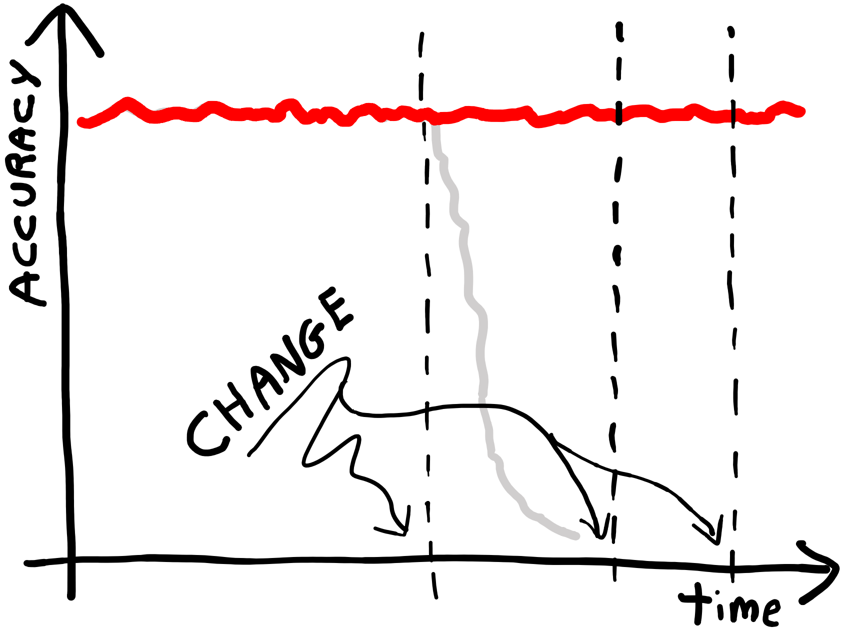
	i.i.d. dataset	i.i.d. data stream	time-dependent time series	evolving data stream	not i.i.d. data stream
Offline Machine Learning	✔	✘	✘	✘	✘
Incremental Learning	✔	✔	✘	✘	✘
Time Series Analysis	✔	~	✔	✘	✘
Streaming Machine Learning	✔	✔	✘	✔	✘

The most general scenario: **not i.i.d. data stream** does not find a comprehensive solution

Time Evolving Analytics framework

Desiderata of the ideal TEA framework:

- Problem agnostic
- No task boundaries
- Stateful learning
- Graceful forgetting
- Selective remembering
- Adaptive learning
- Learning sequences
- Forecasting alternatives



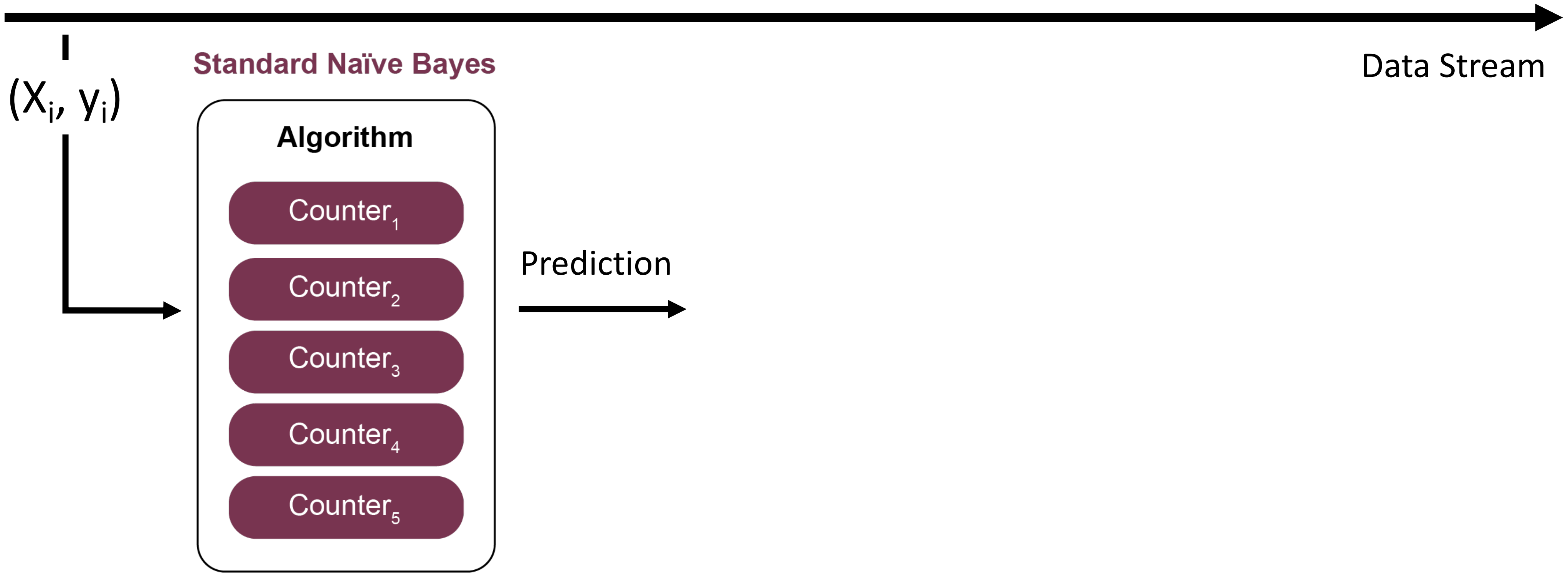
Research question

Aiming at drawing more solid links from data stream learning to the existing time series and sequential learning literature, my work investigated the following research question:

In case of learning from evolving data stream, does the use of online sequential models improve classification with respect to Streaming Machine Learning models?

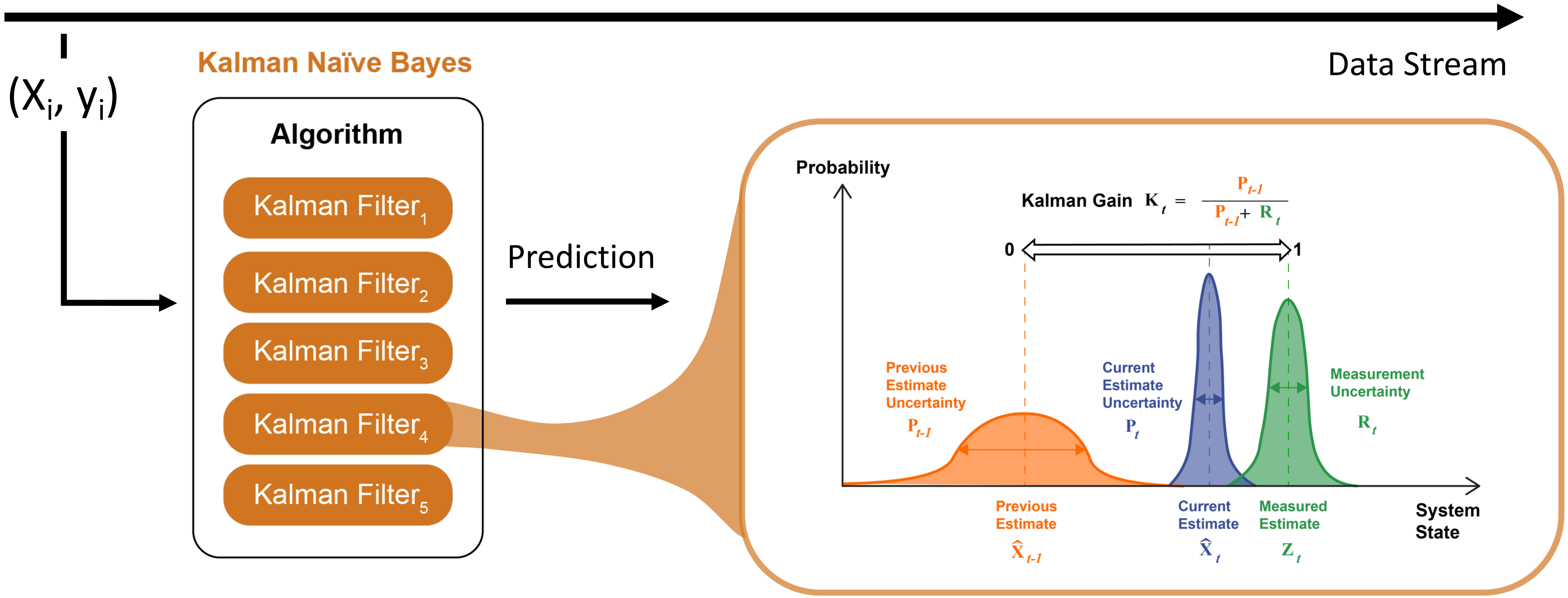
KalmanNB & Hoeffding Kalman Tree

RQ1: *Does the use of a sequential-state space model such as Kalman Filtering helps an online algorithm in learning from evolving data streams?*



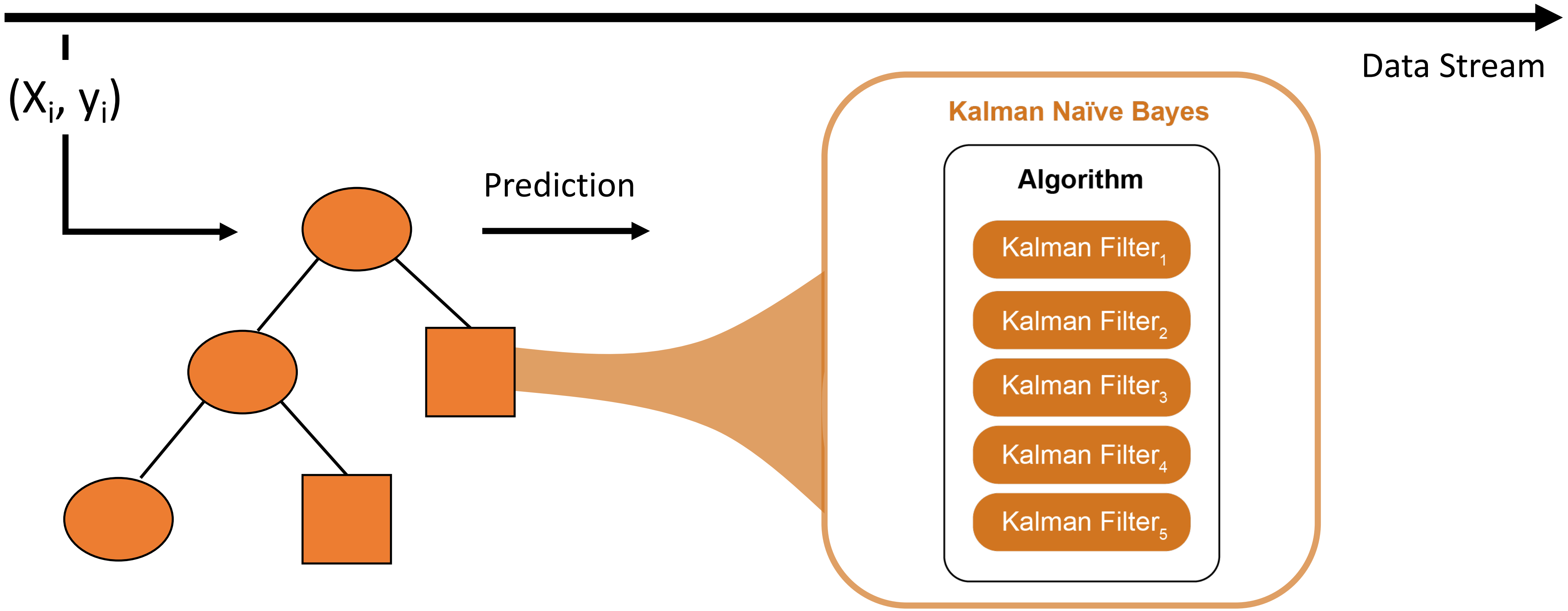
KalmanNB & Hoeffding Kalman Tree

RQ1: *Does the use of a sequential-state space model such as Kalman Filtering helps an online algorithm in learning from evolving data streams?*



KalmanNB & Hoeffding Kalman Tree

Hoeffding Kalman Tree: KalmanNB at the leaves of the Hoeffding Tree to classify data stream



Findings

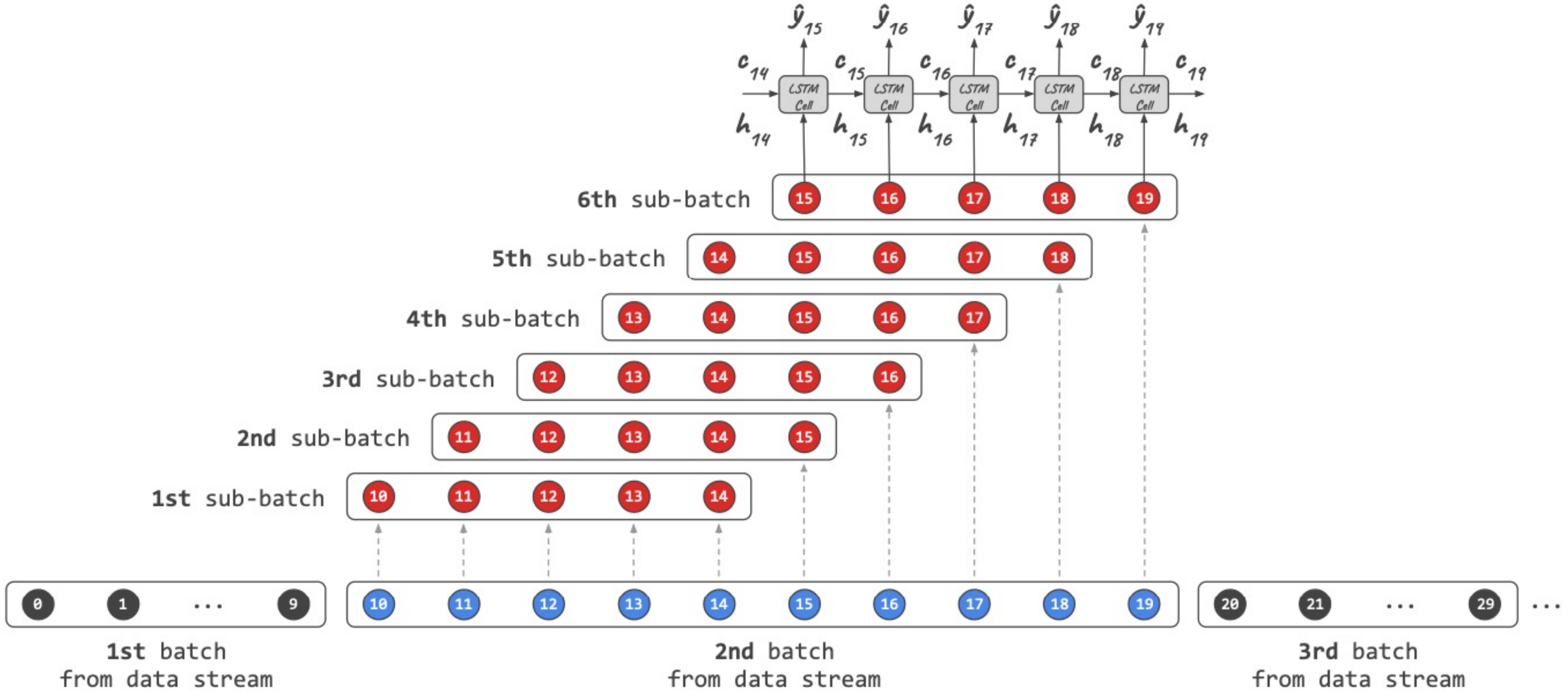
- ⬆ Kalman filtering increase model robustness, making it adaptive to changes at a minor additional computational cost
- ⬆ Processing time and memory consumption are significantly lower than state-of-the-art SML algorithms
- ⬇ Kalman Filtering-based techniques suitable only for categorical features

Ziffer, G., Bernardo, A., Della Valle, E., & Bifet, A. (2021, December). **Kalman Filtering for Learning with Evolving Data Streams**. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 5337-5346). IEEE.

LSTM for time-dependent data streams

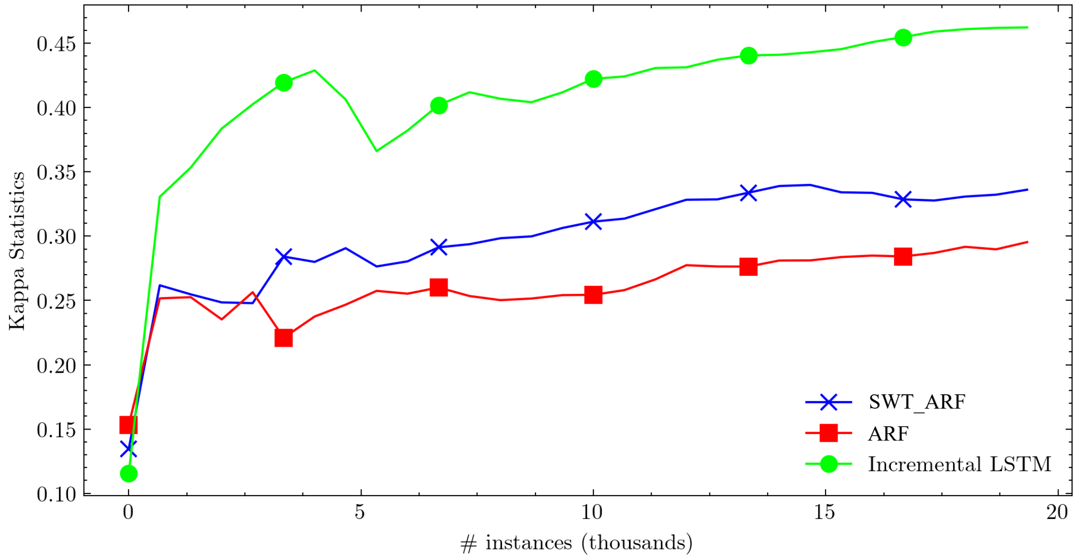
RQ2: *Does the use of an online sequential model, such as Incremental LSTM, improve classification while learning from time-dependent evolving data stream?*

Data stream sub-batching + Incremental LSTM



Findings

⬆ Incremental LSTM proved effective in streaming scenarios where the temporal dependence is relevant



⬇ Not optimal solution to learn from a data stream: high computational resources, i.e., time and RAM, consumption

Ongoing and Future Works

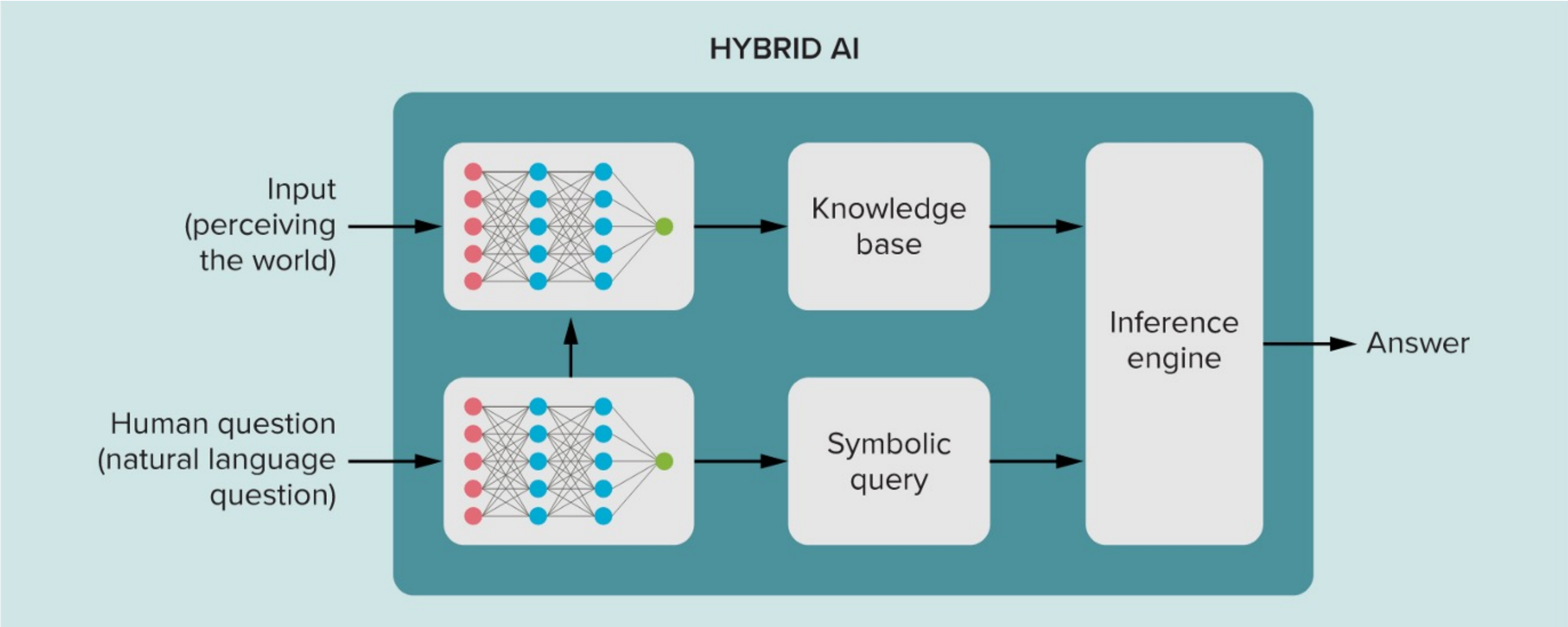
Definition of learning models for time-dependent data streams

- Echo State Networks and Reservoir Computing for stream learning
- Relationships between SML and Continual Learning

Application to heterogenous data

- Definition of high order temporal data stream
- Extension of the methods to learn from unstructured data streams

Time-Evolving Analytics with Symbolic AI



Source: Knowable Magazine

Thank you!

Questions?

Giacomo Ziffer

giacomo.ziffer@polimi.it

gziffer