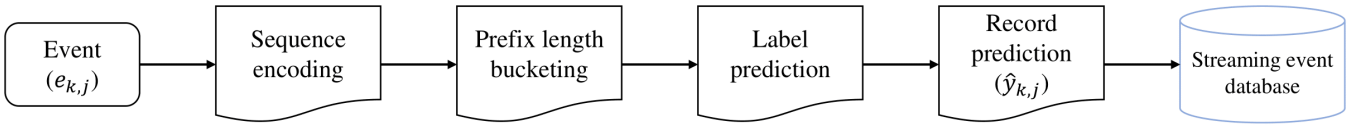
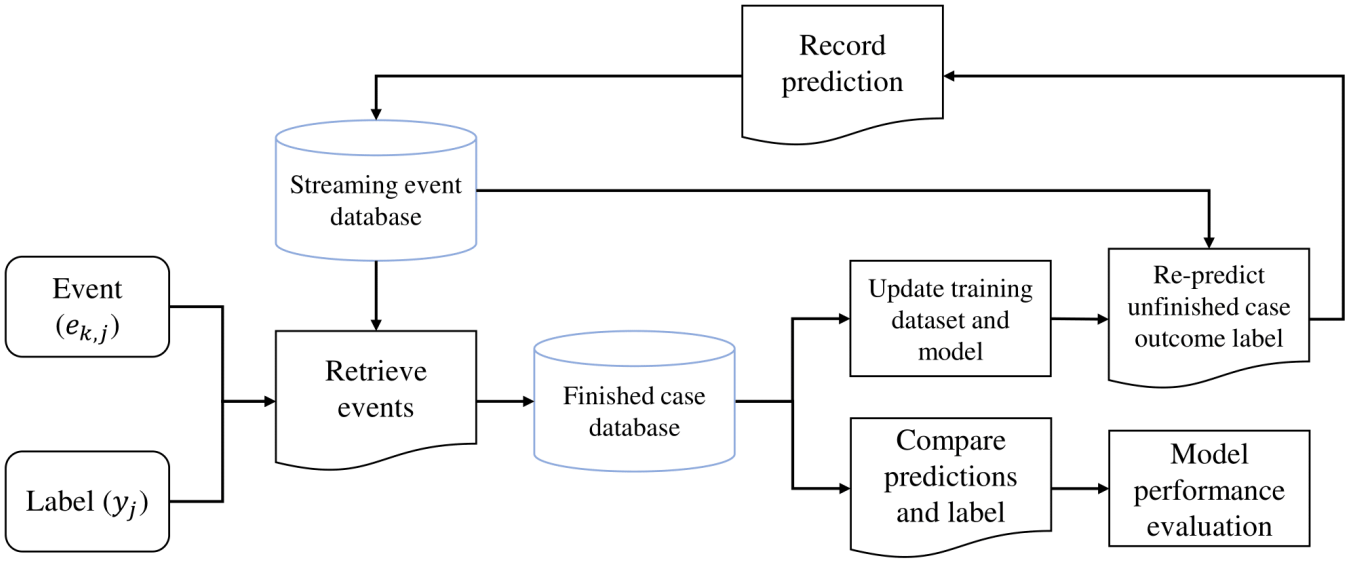
A framework for online business process outcome predictive monitoring

General framework



1.a) Processing of an event without label

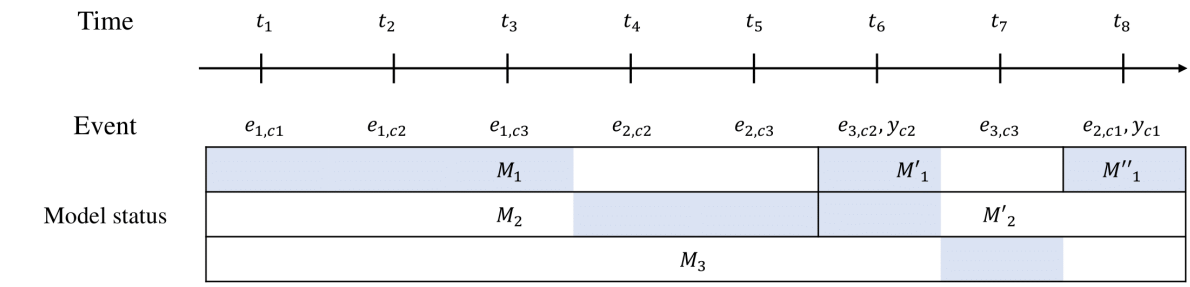


1.b) Processing of an event with label

In this figure, the processing of one event $e_{k,j}$ belonging to trace σ_j is schematised.

First of all, the labels received are only used to train the models in the framework. The event $e_{k,j}$ may either be the last of σ_j, i.e., length of j, in which case the label y_j becomes known, or not. When an event is not the last one of its trace (see Fig. 1.a), it is used to generate a new prefix of σ_j for prefix k. Then, a prediction $\hat{y}*{k,j}$ for the new prefix can be computed using the model for prefix k. Receiving the last event $e*{k,j}$, and its label (see Fig. 1.b) enables (i) to evaluate all the predictions $\hat{y}*j$ that have been generated for all prefixes σ_j (evaluation before training), (ii) to update the models pom_n owing to the availability of new labelled prefixes. Finally, it is possible (iii) to compute a new set of predicted labels $\hat{y}*{k,l}$, with $l\neq j$ and for all the prefixes for which a label has not been yet received (train and retest).

Sample scenario and performance indicator



Label prediction		Timeline								
Case ID	Prefix length	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	$y(\sigma_j)$
1	1	$\hat{y}_{1,c1}$ (0)	>>	>>	>>	>>	$\hat{y}'_{1,c1}$ (0)	>>	-	1
2	1		$\hat{y}_{1,c2}$ (0)	>>	>>	>>	-	-	-	0
	2				$\hat{y}_{2,c2}$ (1)	>>	-	-	-	
3	1			$\hat{y}_{1,c3}$ (0)	>>	>>	$\hat{y}'_{1,c3}$ (1)	>>	$\hat{y}''_{1,c3}$ (1)	Unknown
	2					$\hat{y}_{2,c3}$ (1)	$\hat{y}'_{2,c3}$ (1)	>>	>>	
	3							$\hat{y}_{3,c3}$ (0)	>>	

2) Evaluation methods: supporting example

Fig. 2 exemplifies what stated above in the context of the proposed framework, considering 3 process cases and prefix length up to 3.

Note that different versions of the same model pomk are generated along the considered timeline. In particular, a new version of pom_k is generated when a new label y_j for a case σ_j. For instance, in the example 3 different versions of pom_1 are generated. Second, new predictions for prefixes of length k are generated each time a new version of pom_k is available. In the example, receiving the label y_{c1} of case c1 at t8 triggers (i) the creation of a new version of pom_1 ($pom''*1$) and (ii) the generation of a new prediction $\hat{y}*{1,c3}$ for case c3, which is still running at t8. Finally, note that a prediction can only be evaluated when the corresponding label becomes available. In the example, the predictions generated for all the prefixes of case 3 cannot be evaluated because the label of case 3 has yet to be received at t8.

```
streaming_event_prediction4pm
 ├ data
 │ ├ bpic15.csv
 │ ├ bpic17.csv
 │ ├ iro5k.csv
 │ └ road_traffic_fine_process.csv
 ├ img
 │ └ readme_img
 │ │ ├ Performance_indiciator_figure.png
 │ │ ├ general_framework_a.png
 │ │ └ general_framework_b.png
 ├ result
 │ ├ bpic15
 │ ├ bpic17
 │ ├ iro5k
 │ └ road_traffic_fine_process
```

```
├ .gitignore
├ Readme.md
├ dataset_parameters.json
├ encoding.py
├ offline.ipynb
├ streaming_classification.ipynb
└ utils.py
```