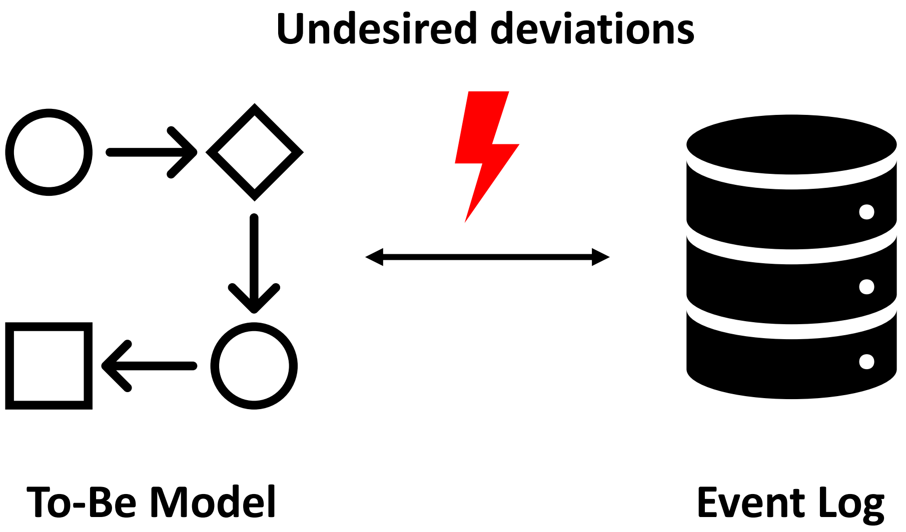
Business Process Deviation Prediction

Predicting Non-Conforming Process Behavior

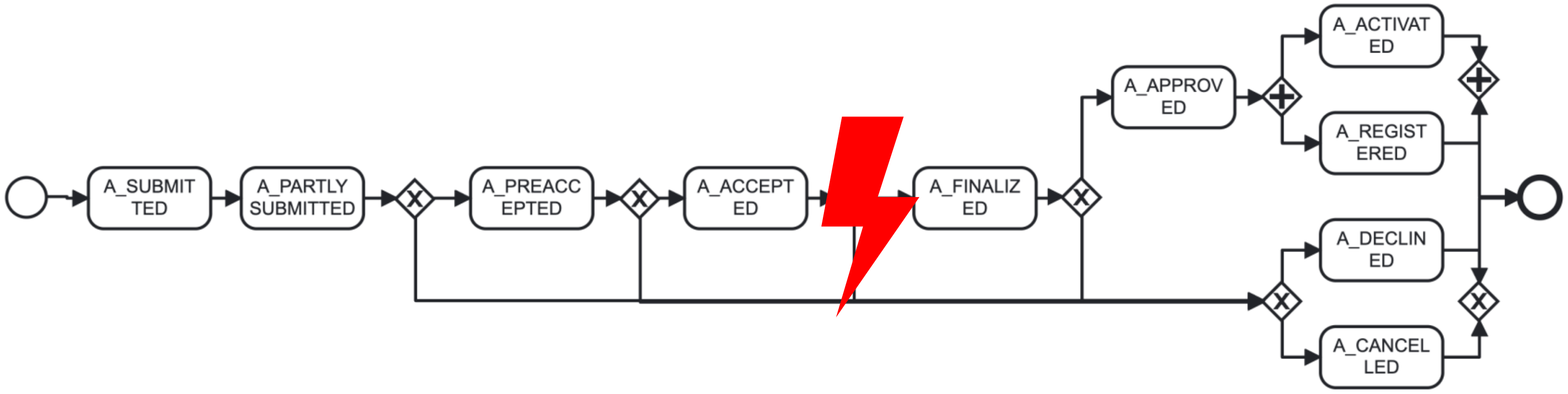
Michael Grohs, Peter Pfeiffer, Jana-Rebecca Rehse

General Setting



Running Example

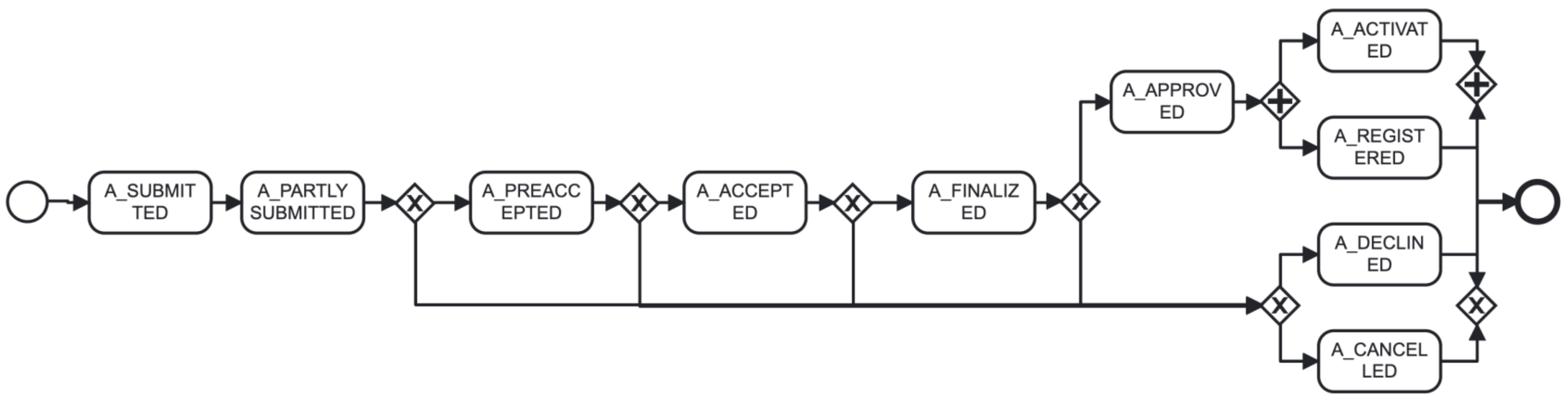
To-Be Model: BPIC 12 only "A_" activities



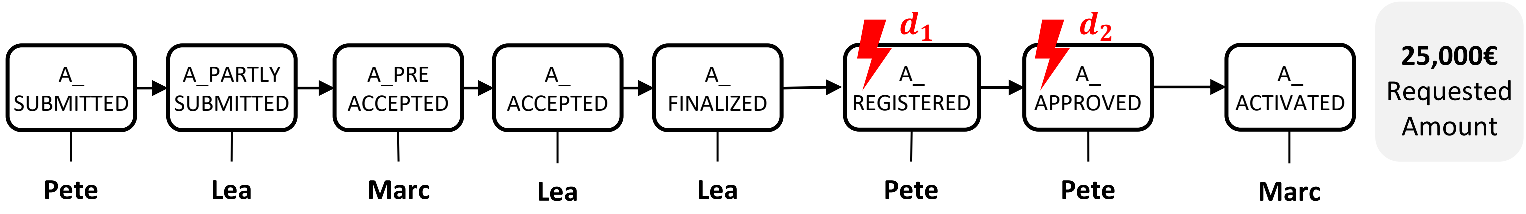
Managers want to know **in which traces which deviations can be expected** and accordingly introduce **preventive measures** → **Our Approach**

Running Example

To-Be Model: BPIC 12 only "A_" activities



Complete Trace t_1



Running Example

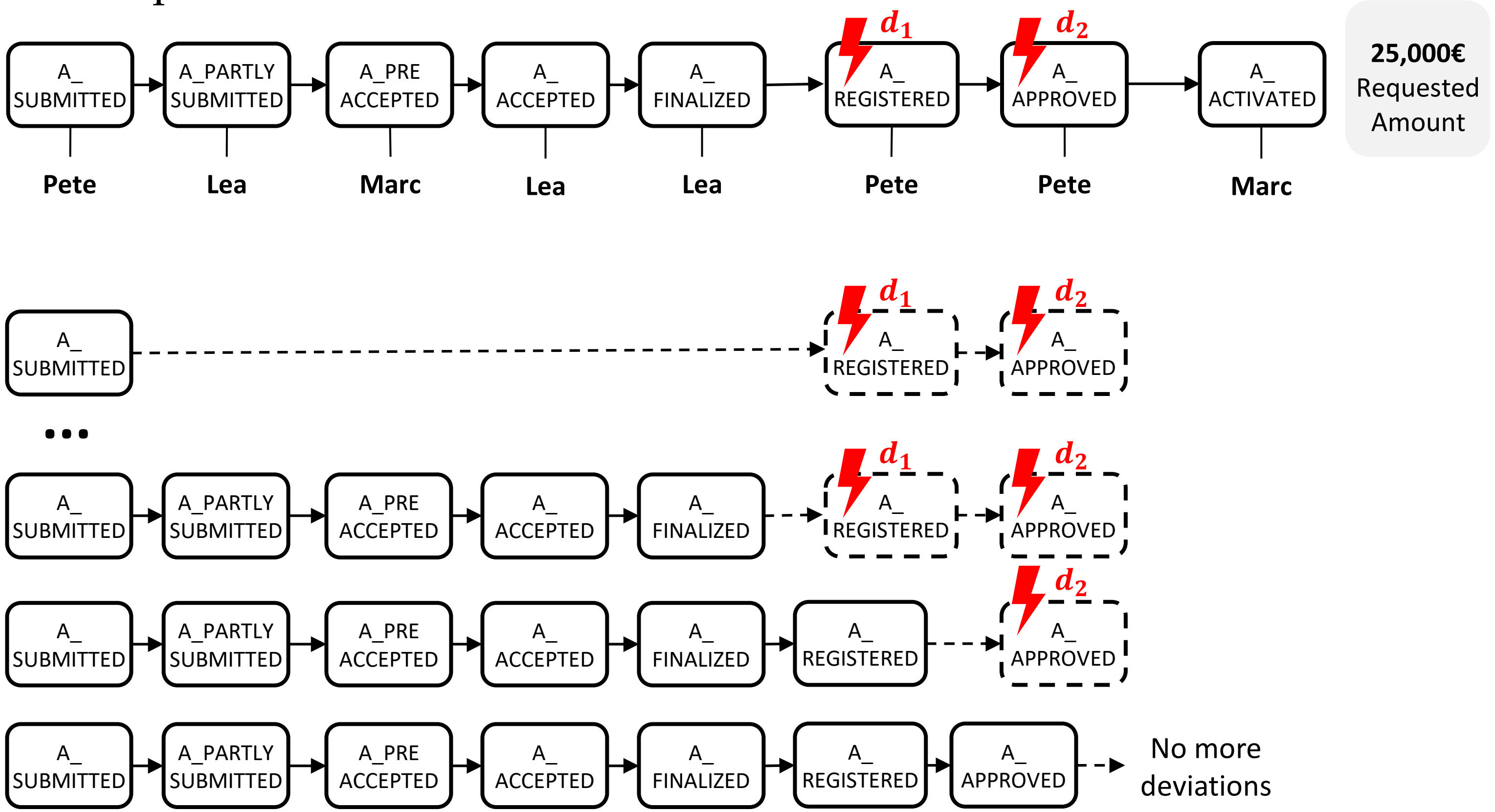
To-Be Model: BPIC 12 only “A_” activities

UNIVERSITY OF MANNHEIM Business School

German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

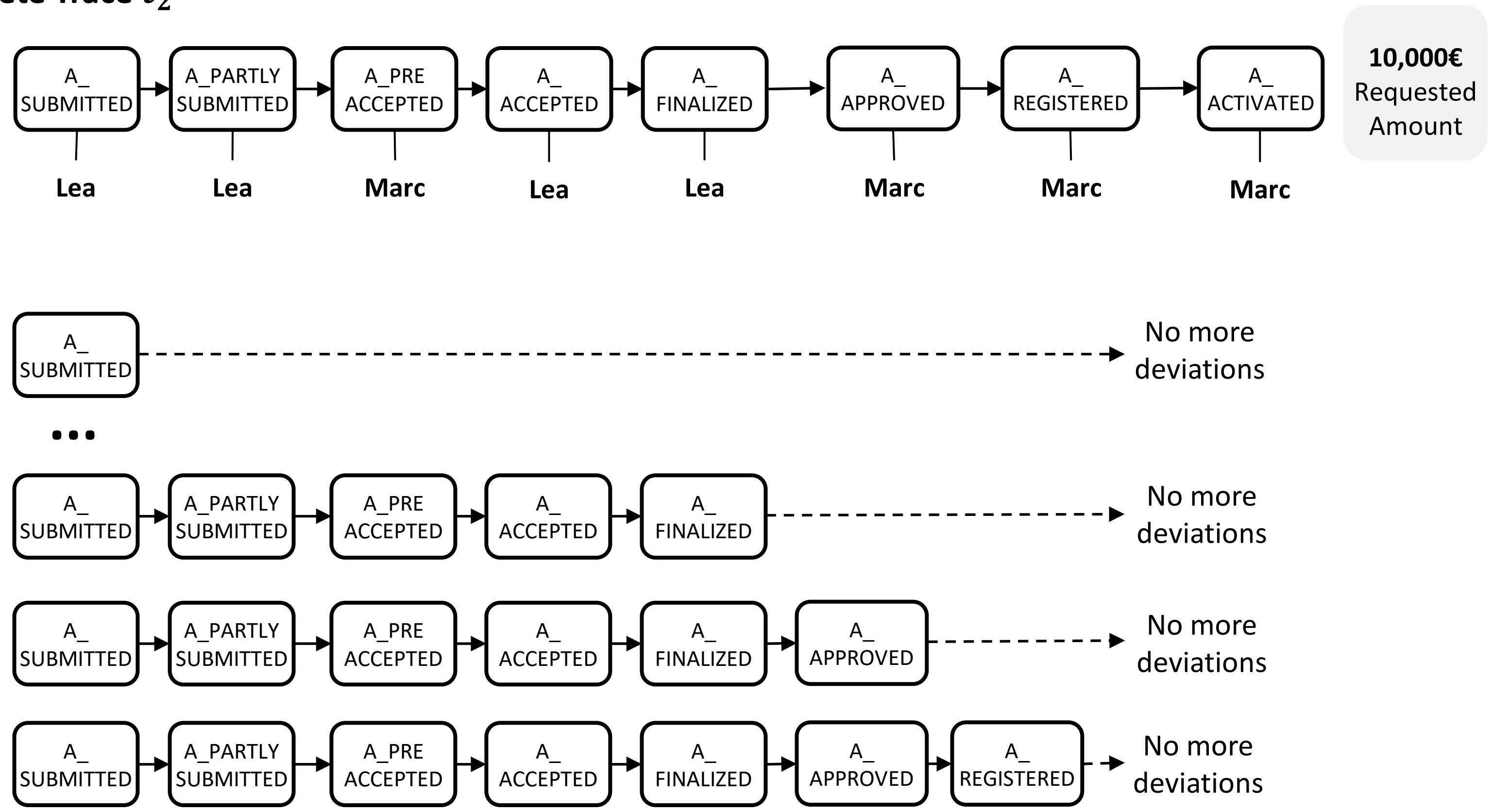
Complete Trace t_1



Running Example

To-Be Model: BPIC 12 only “A_” activities

Complete Trace t_2



Challenges for Deviation Prediction

C_1: Explicit process knowledge required

Prediction targets are clearly defined (unlike in, e.g., anomaly detections)

↳ **usage of to-be model necessary**

Challenges for Deviation Prediction

C_2: Prediction targets have specific nature

$C_{2.1}$: Multi-label targets

one trace might deviate in more than one way (e.g., $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$ in $\boldsymbol{t_1}$)

$C_{2.2}$: Imbalanced targets

deviating traces are rather infrequent, leading to highly imbalanced data

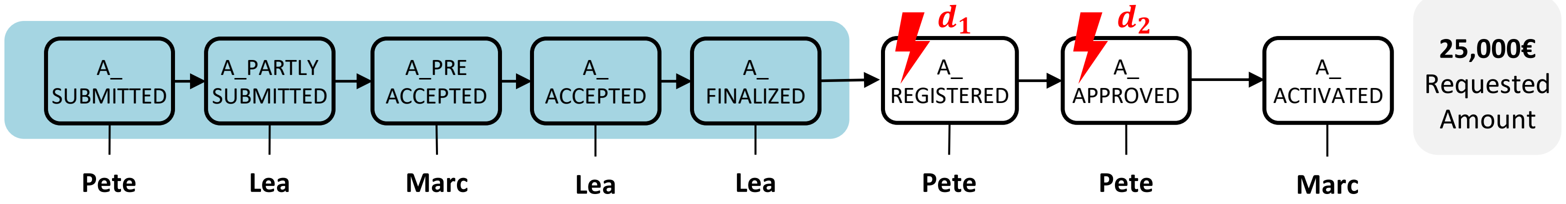
$C_{2.3}$: Dynamic targets

labels change over the duration of the trace (i.e., after deviation happened)

Challenges for Deviation Prediction

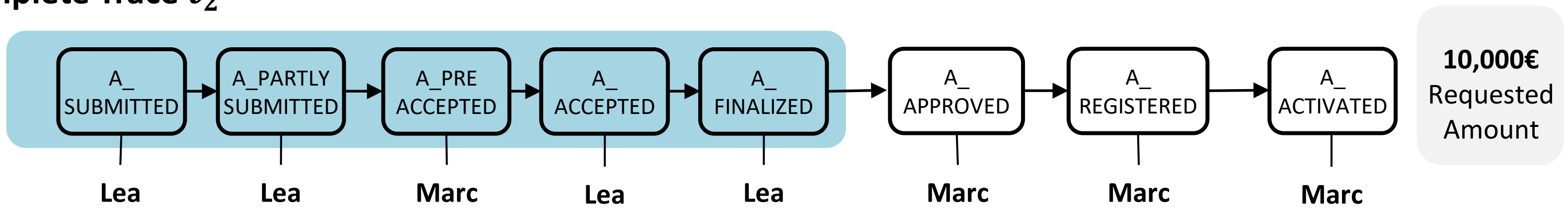
C_3: Context is of importance

Complete Trace t_1



vs.

Complete Trace t_2



Challenges for Deviation Prediction

C_4: Action orientation requires specific focus

Process Manager

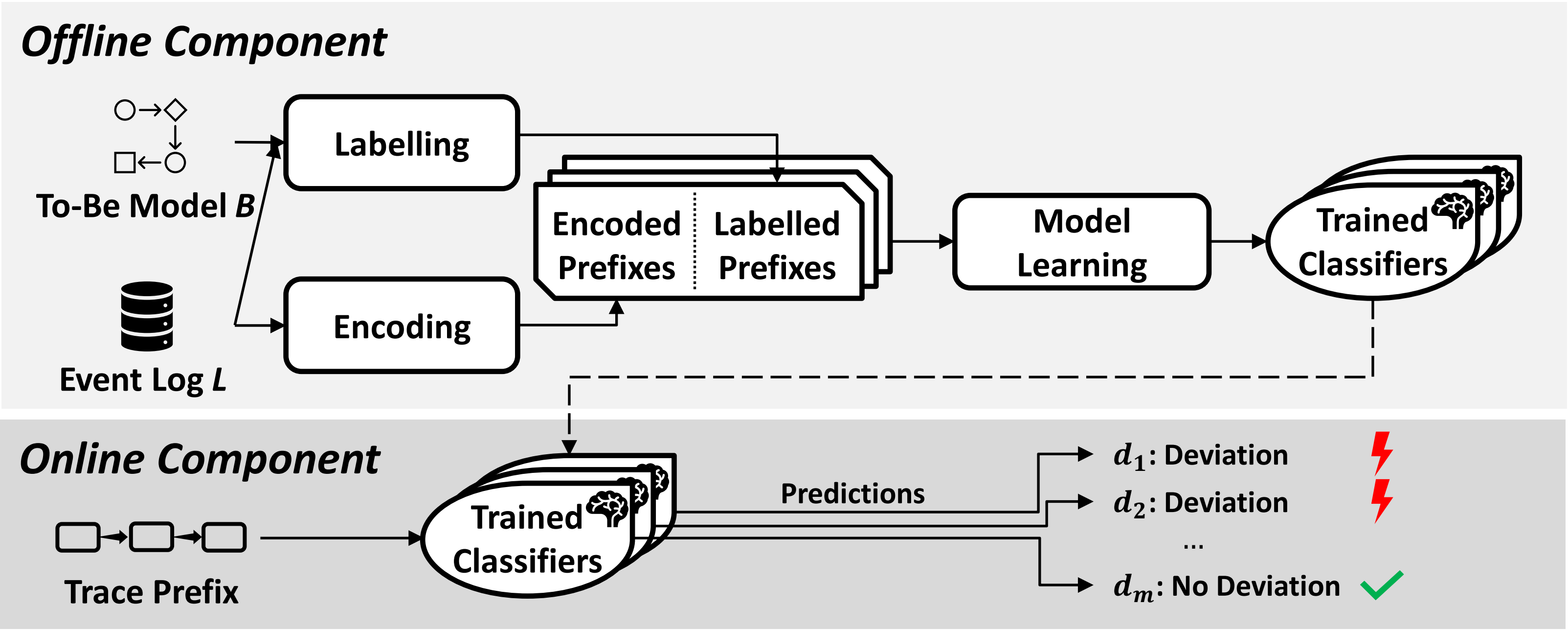
Want **high recall** (correctly recognizing all deviations) **over high precision** (not misclassifying conforming instances)

Default Classifiers
Optimize accuracy

Business Process Deviation Prediction
Overview

Goal

Predict which deviations will happen in the **future of incomplete traces**



Business Process Deviation Prediction
Labelling

Define labels

To consider explicit process knowledge, we define all labels based on **alignments** of traces with to-be models

Addresses C_1

Set of deviation types D

Assign labels

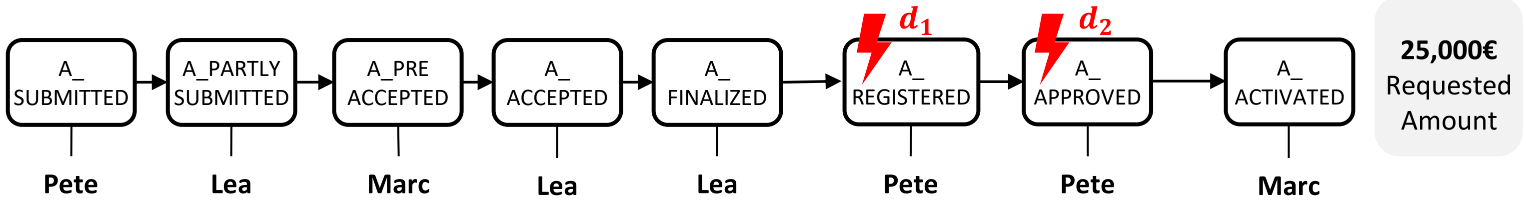
To account for multiple targets that change dynamically throughout the trace, we **label each prefix individual**

Addresses $C_{2.1}$ & $C_{2.3}$

Label for each prefix for each $d \in D$

Business Process Deviation Prediction
Labelling - Example

Complete Trace t_1



Corresponding Labels

SUB = A_SUBMITTED; PAR = A_PARTLY SUBMITTED;
PRE = A_PREACCEPTED; ACC = A_ACCEPTED; FIN = A_FINALIZED;
REG = A_REGISTERED; APP = A_APPROVED; ACT = A_ACTIVATED

Trace Prefix	Labels	
	$d_1 : (\gg, \text{APP})$	$d_2 : (\text{APP}, \gg)$
⟨SUB⟩	1	1
⟨SUB, PAR⟩	1	1
⟨SUB, PAR, PRE⟩	1	1
⟨SUB, PAR, PRE, ACC⟩	1	1
⟨SUB, PAR, PRE, ACC, FIN⟩	1	1
⟨SUB, PAR, PRE, ACC, FIN, REG⟩	0	1
⟨SUB, PAR, PRE, ACC, FIN, REG, APP⟩	0	0
⟨SUB, PAR, PRE, ACC, FIN, REG, APP, ACT⟩	0	0

Business Process Deviation Prediction

Encoding

Apply and compare two context-aware encodings

Complex Index-Based Encoding (CIBE)

Learned Encoding (MPPN)

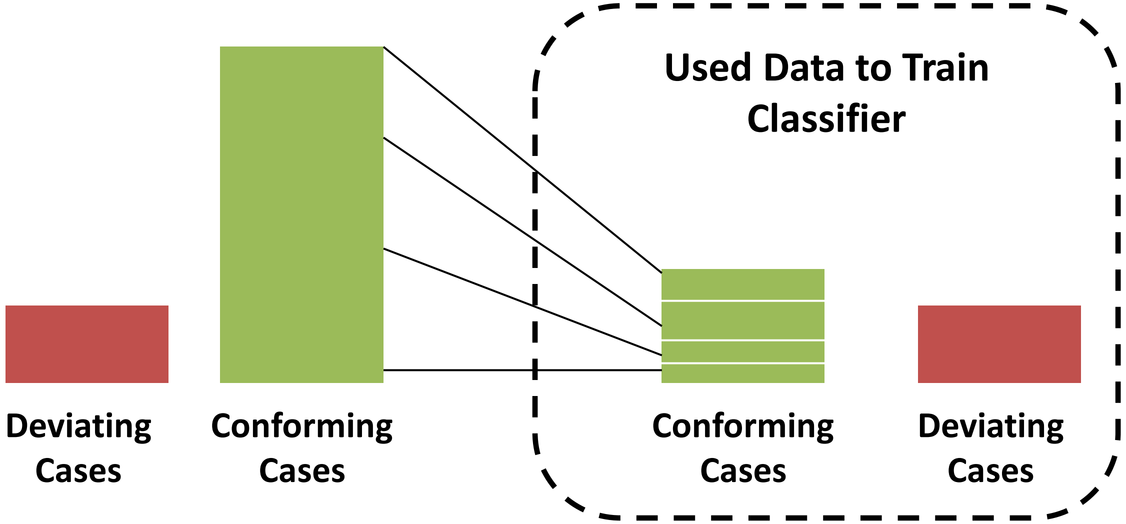
Addresses $\boldsymbol{C_3}$

Business Process Deviation Prediction

Learning - Sampling

We apply under-sampling to the training split

Addresses
$C_{2.2}$



Applied Under-Sampling:
One-Sided Selection

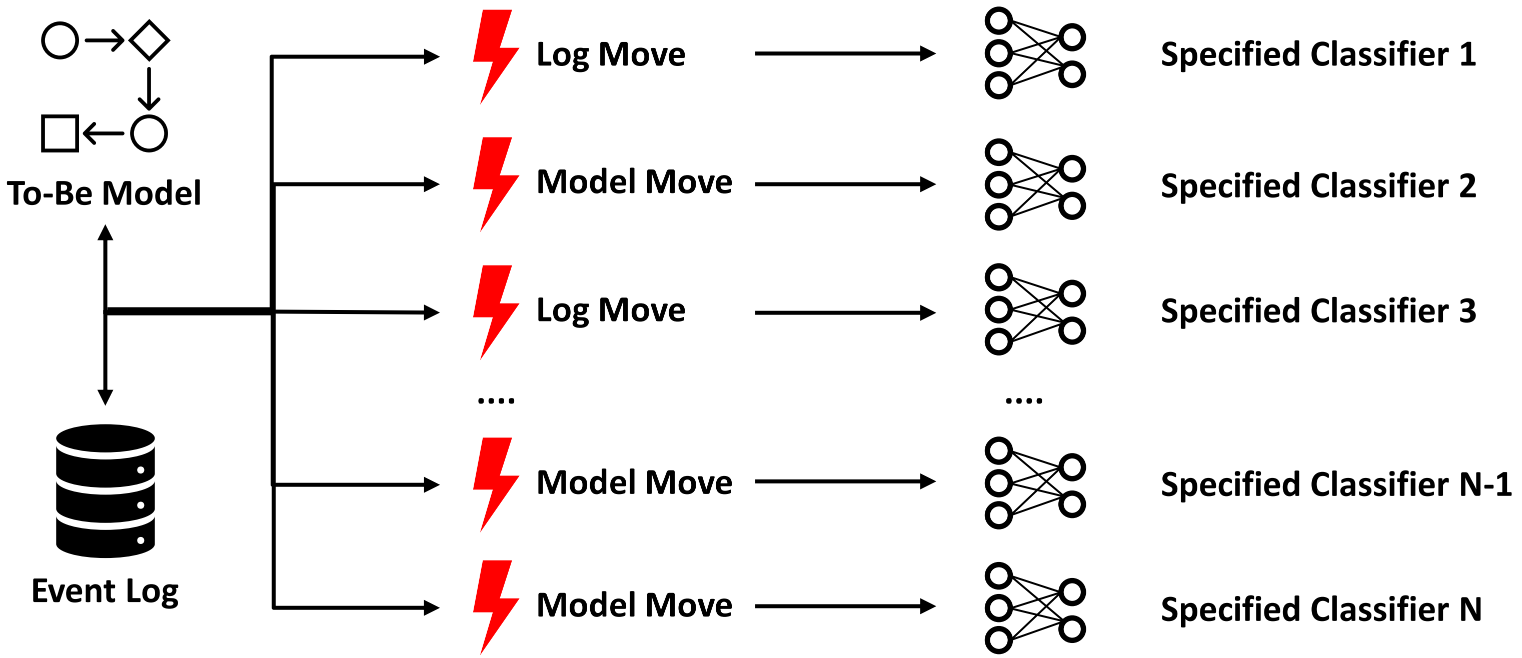
- Combines **Tomek Links and Condensed Nearest Neighbor (CNN) Rule**
- Tomek Links removes ambiguous samples
- CNN removes redundant samples

Business Process Deviation Prediction

Learning - Network Architecture

We train one classifier per $d \in D$

Addresses $C_{2.1}$



Business Process Deviation Prediction

Learning - Training

We use a weighted cross-entropy loss function $(WCEL)$

Addresses C_4

$$loss_{WCEL} = \begin{cases} 16 \; if \; False \; Negative \\ 1 \; if \; False \; Positive \end{cases}$$

✓ **Penalizes FN more**

✓ **Optimizes Recall**

Evaluation

Datasets

Log L		Traces	Events	Trace Attr.	Trace Length min.	Trace Length avg.	Trace Length max.	Dev. Types	Deviating Traces min.	Deviating Traces avg.	Deviating Traces max.
BPIC 12	A	13,087	60,849	1	3	4.7	8	3	399	927	1,191
	O	5,015	31,244	1	3	6.2	30	8	20	984	1,761
BPIC 20	Dom. Dec.	10,500	56,437	4	1	5.4	24	19	1	252	2,154
	Int. Dec.	6,449	72,151	17	3	11.2	27	46	1	292	1,701
	RfP	6,886	36,796	8	1	5.3	20	23	1	116	1,027
	Prep.	2,099	18,246	16	1	8.7	21	41	1	64	530
MobIS		3,354	55,809	1	11	16.6	49	43	1	182	1011

Differing numbers of prefixes

High imbalance of deviations

Differing amount of context

Many deviation types $d \in D$

Evaluation

Results

Log	Baselines			BPDP	
	Genga et. al.	CatBoost	Suffix Prediction	$BPDP_{CIBE}$	$BPDP_{MPPN}$

Approach with similar goal based on statistics

Approaches that could be used instead of BPDP

BPDP with both CIBE and MPPN encodings

Evaluation

Results

Log		Baselines						BPDP					
		Genga et. al.		CatBoost		Suffix Prediction		$BPDP_{CIBE}$		$BPDP_{MPPN}$	
		Dev	No Dev	Dev	No Dev	Dev	No Dev	Dev	No Dev	Dev	No Dev
BPIC 12A	Precision										
	Recall										
	AUC_{ROC}										
BPIC 12O	Precision										
	Recall										
	AUC_{ROC}										
BPIC 20 Dom. Dec.	Precision										
	Recall										
	AUC_{ROC}										
BPIC 20 Int. Dec.	Precision										
	Recall										
	AUC_{ROC}										
BPIC 20 RfP	Precision										
	Recall										
	AUC_{ROC}										
BPIC 20 Prep.	Precision										
	Recall										
	AUC_{ROC}										
MobIS	Precision										
	Recall										
	AUC_{ROC}										

Evaluation
Results

are outperformed

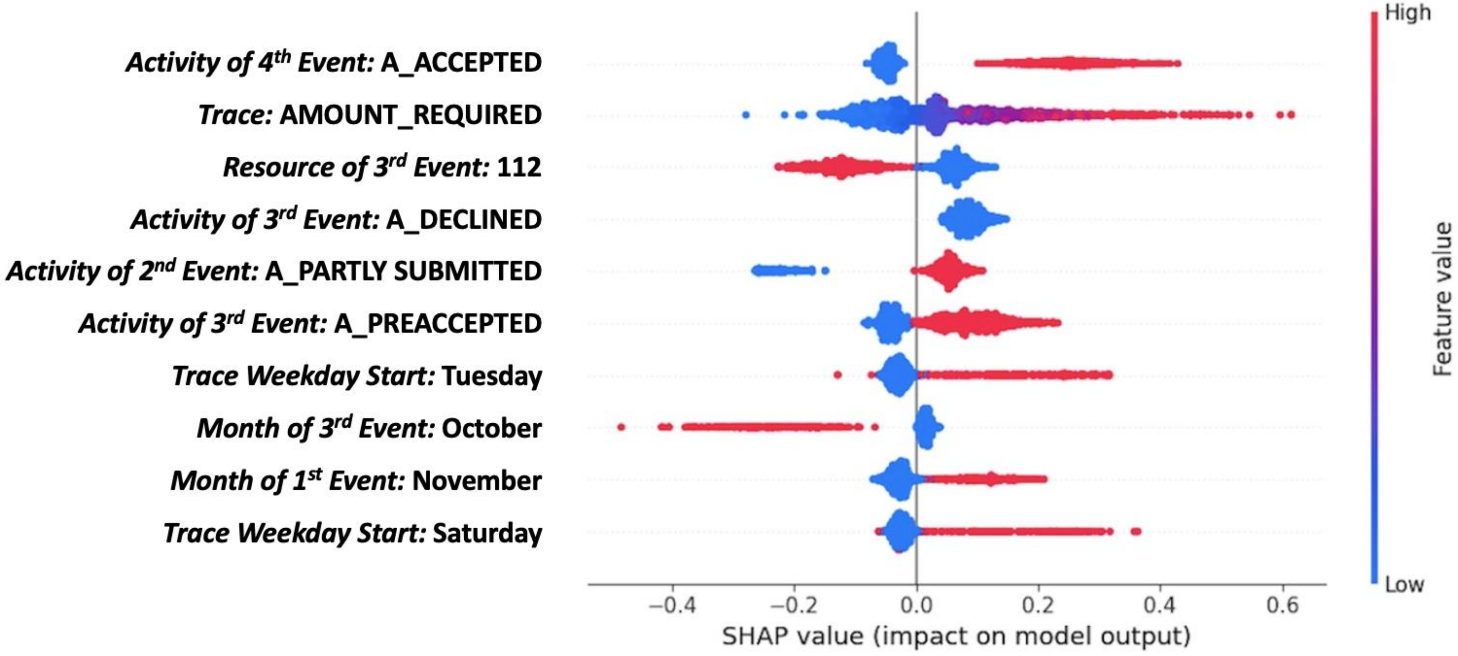
best performance

Log		Baselines						BPDP			
		Genga et. al.		CatBoost		Suffix Prediction		$BPDP_{CIBE}$		$BPDP_{MPPN}$	
		Dev	No Dev	Dev	No Dev	Dev	No Dev	Dev	No Dev	Dev	No Dev
BPIC 12A	Precision	0.2136	0.9214	0.1694	0.9622	0.3766	0.9256	0.1620	0.9669	0.1405	0.9480
	Recall	0.0678	0.9654	0.7110	0.6797	0.0967	0.9895	**0.8084**	0.6563	0.6636	0.5733
	AUC_{ROC}	0.5166		0.6954		0.5431		**0.7324**		0.6185	
BPIC 12O	Precision	0.1462	0.8663	0.2277	0.9566	0.3282	0.8811	0.2019	0.9625	0.1798	0.9652
	Recall	0.1340	0.8403	0.6034	0.7148	0.2128	0.8504	0.7981	0.5348	**0.8089**	0.4588
	AUC_{ROC}	0.4872		0.6591		0.5316		**0.6665**		0.6339	
BPIC 20 Dom. Dec.	Precision	0.1961	0.7318	0.4035	0.9977	0.1934	0.9964	0.1401	0.9982	0.0314	0.9980
	Recall	0.2040	0.7205	0.5110	0.9930	0.2900	0.9868	**0.7619**	0.8897	0.6459	0.7876
	AUC_{ROC}	0.6372		0.7511		0.7023		**0.8258**		0.7168	
BPIC 20 Int. Dec.	Precision	0.1738	0.8823	0.3648	0.9955	0.1096	0.9911	0.0720	0.9938	0.0741	0.9973
	Recall	0.1648	0.8684	0.3866	0.9884	0.2652	0.9732	**0.6333**	0.8223	0.6239	0.8456
	AUC_{ROC}	0.5796		0.6969		0.6338		0.7270		**0.7348**	
BPIC 20 RfP	Precision	0.1480	0.7352	0.4630	0.9979	0.2259	0.9972	0.0402	0.9979	0.0291	0.9985
	Recall	0.1244	0.7273	0.3967	0.9965	0.2064	0.9908	**0.6888**	0.8353	0.6486	0.8180
	AUC_{ROC}	0.5762		0.6961		0.6335		**0.7620**		0.7333	
BPIC 20 Prep.	Precision	0.1475	0.8705	0.3032	0.9949	0.1373	0.9943	0.0457	0.9969	0.0270	0.9971
	Recall	0.1067	0.8629	0.2727	0.9946	0.2447	0.9804	**0.5566**	0.8514	0.5511	0.7664
	AUC_{ROC}	0.5521		0.6333		0.6284		**0.7040**		0.6587	
MobIS	Precision	0.1211	0.8355	0.1363	0.9296	0.1176	0.9697	0.0993	0.9748	0.0956	0.9971
	Recall	0.1245	0.8415	0.2254	0.8907	0.2063	0.9599	**0.7162**	0.5906	0.5644	0.7391
	AUC_{ROC}	0.5461		0.5729		0.5958		**0.6534**		0.6518	

Evaluation

Shapley Values for one specific prediction

Applying XAI to identify which features lead to a deviation prediction



unlikely to predict a deviation → **likely** to predict a deviation

Discussion

Further evaluation results

- **Shapley values** allow process managers to **detect features that increase likelihood of deviations**
- Additional use case evaluation showed that **BPDP is early in its predictions**

Insights into deviation prediction

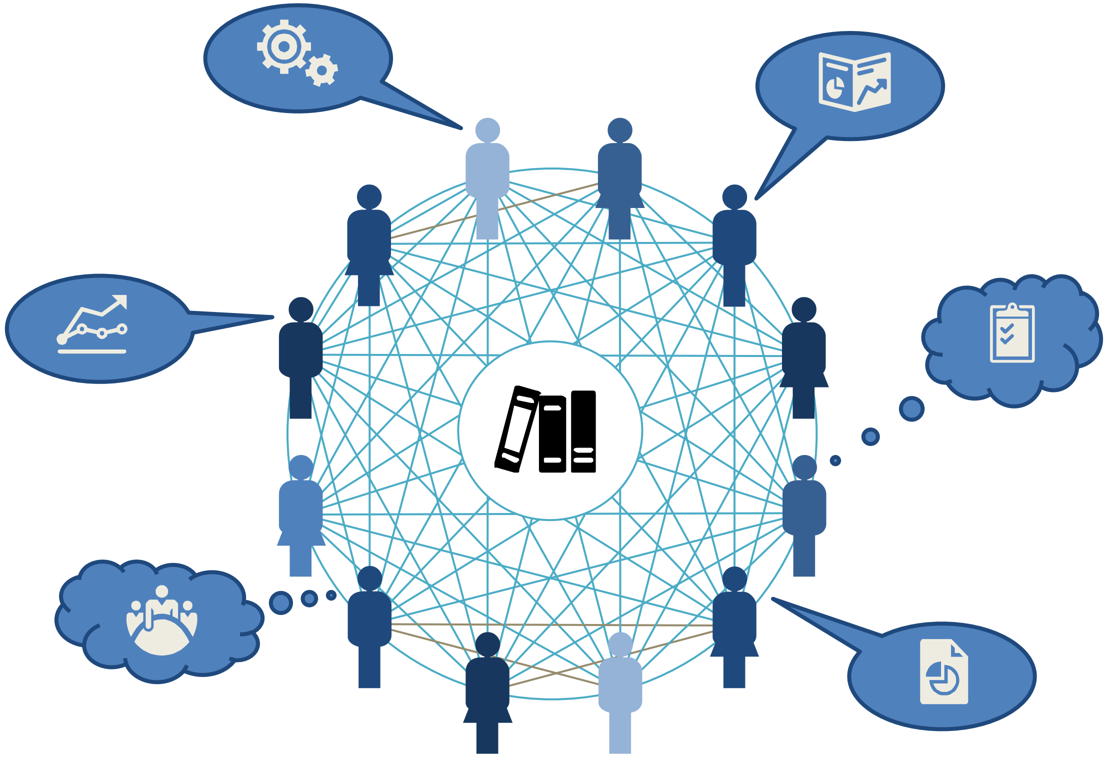
- Only **both under-sampling and WCEL** leads to sufficient results
- Learned encoding **MPPN** performs **nearly as good as CIBE** although trained on next-event prediction
- **Additional context** in BPIC 20 logs could lead to **better performance for shorter prefixes**

Any Questions?

UNIVERSITY OF MANNHEIM
Business School

German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

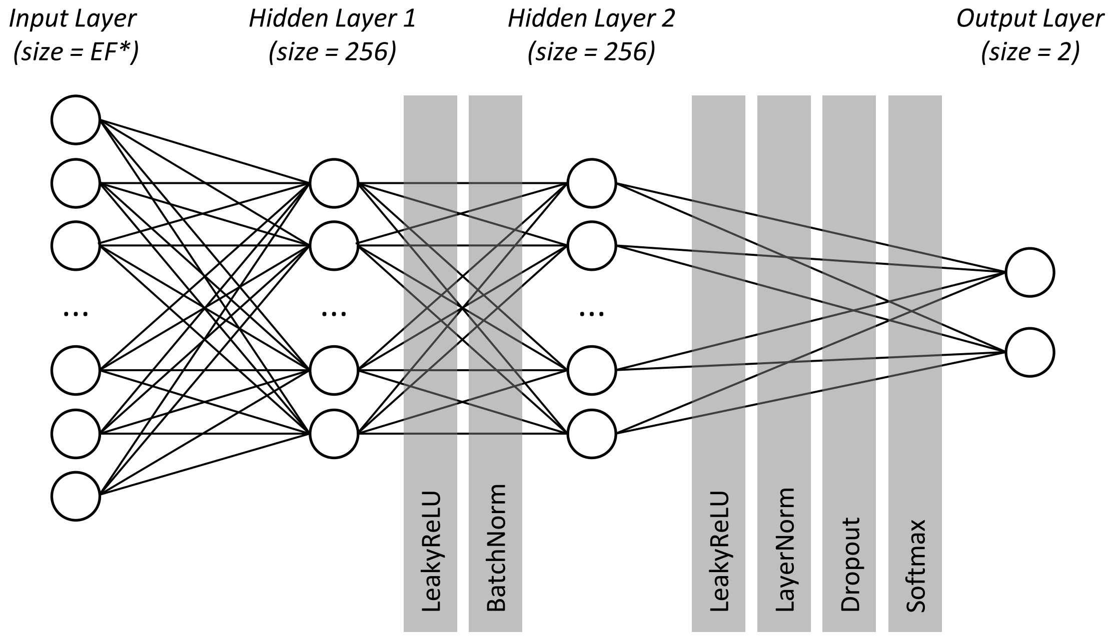


Backup

Business Process Deviation Prediction

Learning - Network Architecture

We train one classifier per $d \in D$



**EF = number of encoded features for each trace prefix*