

In this thesis, we use ten publicly available datasets. All of the datasets were taken from Teinemaa *et al.* Each dataset consists of log data and contains labels that signify a process’s outcome. We focus on binary outcome predictions. Hence, each dataset provides a binary label for each process instance that indicates the outcome of that process instance.

**BPIC12:** The first dataset is the popular BPIC12 dataset. This dataset was originally published for the Business Process Intelligence Conference and contains events for a loan application process. Each case relates to one loan application process and can be accepted (regular) or cancelled (deviant).

**Sepsis:** The next dataset is the Sepsis-Dataset. It is a medical dataset that records patients with life-threatening sepsis conditions. The outcome describes whether the patient returns to the emergency room within 28 days from initial discharge.

**TrafficFines:** Third, we apply our approach to the Traffic-Fines-Dataset. This dataset contains events related to notifications sent related to a fine. The dataset originates in a log from an Italian local police force.

**DiCE4EL:** Lastly, we include a variation of the BPIC dataset. It is the dataset which was used by Hsieh, Moreira, and Ouyang. The difference between this dataset and the original dataset is two-fold. First, Hsieh, Moreira, and Ouyang omit most variables except two. Second, it is primarily designed for next-activity prediction and not outcome prediction. We modified the dataset to fit the outcome prediction model.

Dataset	#Cases	Min Len	Max Len	% Unique Traces	#Unique Ev.	#Data Columns	#Event Attr	#Regular	#Deviant
DiCE4EL	3 051	12	25	0.000328	23	9	7	1 853	1 198
BPIC12-25	3 051	12	25	0.000328	23	23	21	1 853	1 198
BPIC12-50	4 587	12	50	0.000218	23	23	21	2 405	2 182
BPIC12-75	4 677	12	75	0.000214	23	23	21	2 436	2 241
BPIC12-100	4 685	12	96	0.000213	23	23	21	2 442	2 243
Sepsis-25	707	5	25	0.001414	15	75	73	610	97
Sepsis-50	770	5	47	0.001299	15	76	74	662	108
Sepsis-75	777	5	66	0.001287	15	76	74	667	110
Sepsis-100	779	5	88	0.001284	15	76	74	669	110
TrafficFines	129 615	2	20	0.000008	10	40	38	70 602	59 013

Table 1: All datasets used within the evaluation. DiCE4EL is used for the qualitative evaluation, and the remaining are used for quantitative evaluation purposes.

For more information about these datasets we refer to Teinemaa *et al.*’s comparative study[2]. We list all the important descriptive statistics in Table 1.