In this thesis, we use a multitude of datasets for generating the counterfactuals. All of the data sets where taken from Teinemaa et al. Each dataset consists of log data and contains labels which signify the outcome of a process. We focus on binary outcome predictions. Hence, each dataset will provide information about one of two possible outcomes associated with the case. For instance, a medical process might be deemed a success if the patient is cured or a failure if the patient remains ill. A loan application process might deem granting the loan a success or the rejection as failure. The determination of the outcome depends on the use-case and the stakeholders involved. A insurence provider might deem a successful claim as a failure, while the client deems it as a success.

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 indivdual 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, which records of 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 et al. The difference between this dataset and the original dataset is two-fold. First, Hsieh et al. 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.

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.

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	866	15	25	0.001155	32	23	21	682	184
BPIC12-50	3 728	15	50	0.000268	36	25	23	2 111	1 617
BPIC12-75	4 461	15	75	0.000224	36	25	23	2 379	2 082
BPIC12-100	4 628	15	100	0.000216	36	25	23	2 420	2 208
Sepsis25	707	5	25	0.001414	15	75	73	610	97
Sepsis50	770	5	47	0.001299	15	76	74	662	108
Sepsis75	777	5	66	0.001287	15	76	74	667	110
Sepsis100	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.