americanamerican-apa

./references/autoupdated.bib

Many processes, often medical, economical, or administrative in nature, are governed by sequential events and their contextual environment. Many of these events and their order of appearance play a crucial part in the determination of every possible outcomevanderaalst$_{P}rocessMiningManifesto_{2}012.WiththeriseofAIand$

Research in the Process Mining discipline has shown that it is possible to predict the outcome of a particular process fairly welltax$_{P}redictiveBusinessProcess_{2}017a, klimek$

This difficulty arises from models, like neural networks, that are so-called *blackbox models*. Meaning, that their inference is incomprehensible, due to the vast amount of parameters involved. This lack of comprehension is undesirable for many fields like IT or finance. Not knowing why a loan was given, makes it impossible to rule out possible biases. Knowing what will lead to a system failure will help us knowing how to avoid it. In critical domains like medicine, the reasoning behind decisions becomes crucial. For instance, if we know that a treatment process of a patient reduces the chances for survival, we want to know which treatment step is the critical factor we ought to avoid. To summarise, knowing the outcome of a process often leads us to questions on how to change it. Formally, we want to change the outcome of a process instance by making it maximally likely with as little interventions as possiblemolnar2019. **??** is a visual representation of the desired goal.

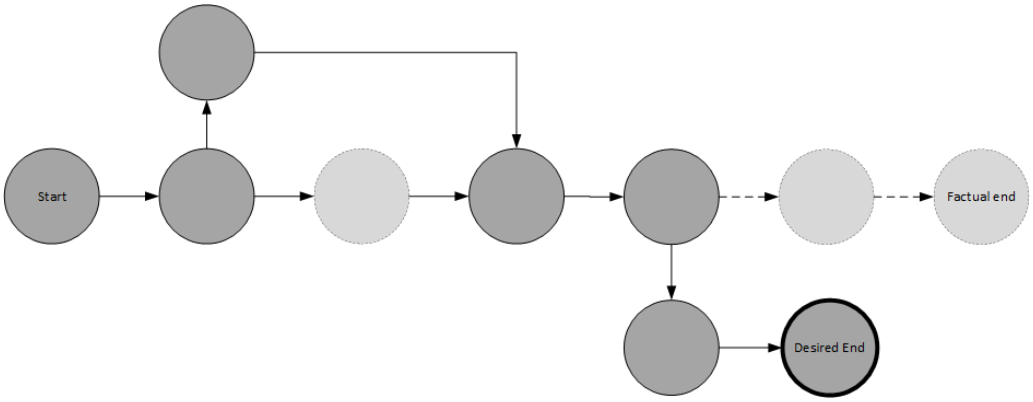

Figure 1: This figure illustrates a model, that predicts a certain trajectory of the process. However, we want to change the process steps in such a way, that it changes the outcome.

One way to better understand the Machine Learning (ML) models lies within the eXplainable AI (XAI) discipline. XAI focuses the developments of theories, methods, and techniques that help explaining blackbox models models to humans. Most of the discipline's techniques produce explanations that guide our understanding. Explanations can come in various forms, such as

IF-THEN rules[p.90]molnar2019 or feature importance[p.45]molnar2019, but some are more comprehensible for humans than others.

A prominent and human-friendly approach are *counterfactuals*[p. 221]molnar2019. Counterfactuals within the AI framework help us to answer hypothetical "what-if" questions. Basically, if we know *what* would happen *if* we changed the execution of a process instance, we could change it for the better. In this thesis, we raise the question how we can use counterfactuals to change the trajectory of a process models' prediction towards a desired outcome. Knowing the answers not only increases the understanding of blackbox models, but also help us avoid or enforce certain outcomes.