Process Miner, are you sure?
On reliability and validity in process mining research

Major credits go to my colleagues
Jan Martijn van der Werf,
Sander Leemans, and Peter Fettke

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
How do you know that your process mining approach still works when...

- it is applied by someone else?
- it is applied in another setting?
- it is applied on new data?

Will your claims about the approach still hold?

Our new (really cool!) process mining approach

1) Introduction
2) Foundations
3) Approach
   a) Step 1
   b) Step 2
   c) Step 3
4) Evaluation
5) Related Work
6) Discussion
7) Conclusion

It provides better results than the other approach!
It works on real-life data!
It can deal with large event logs!
It achieved an F-score above 0.8!
A shift in process mining research

Historic development
• Theoretical computer science, mathematical modeling, automata theory
• Typical results: properties of models and / or algorithms
• Typical methods: (formal) mathematical proofs

Current reality
• Process mining is applied in organizations and practical settings
• Focus lies on data analysis
• Data is influenced by IT systems, people and other social constructs
• Properties of this data cannot be formally proven

Conclusion: We need to use more empirical methods from social science instead of formal methods from mathematics or CS.
Process Discovery: Illustration

Which model represents the process best?

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
How should we weigh the different dimensions?

How can we measure them?

“able to replay event log”

“Occam’s razor”

replay fitness

simplicity

“not overfitting the log”

“not underfitting the log”

generalization

precision

process discovery

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
Process Discovery: Illustration

"True" Process

represents

generates

Process Model

Event Log

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
Process Discovery: Illustration

"True" Process

represents

generates

Process Model

discovers

Can you discover the same process from other logs?

Event Log
Process Discovery: Illustration

Can you discover models of similar quality for other processes?

Can you discover the same process from other logs?

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
Principles of scientific inquiry

Degree to which a measure of a concept is stable

Reliability

Scientific Inquiry

Validity

Degree to which results are free from errors

Unreliable & Not Valid

Unreliable, But Valid

Reliable, Not Valid

Both Reliable & Valid

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
Reliability

Degree to which a measure of a concept is stable

Same researchers, same setup → similar results?

Other researchers, same setup → similar results?

Scientific Inquiry

Reliability

Repeatability

Direct Replicability

Conceptual Replicability

Validity

Replicability

(Other researchers), other setup → similar results?
Validity

Scientific Inquiry

Reliability

Validity

Degree to which results are free from errors

Experiment allows for stated conclusions

Conclusion Validity

Construct Validity

Internal Validity

External Validity

Identified causalities also hold in other settings

Measures assess the intended property

Soundness

Completeness

Observed effects can be attributed to treatment

Prof. Dr. Jana-Rebecca Rehse
ICPM 2023 Doctoral Consortium Keynote
What could possibly go wrong?

Degree to which a measure of a concept is stable

Reliability

Repeatability

Replicability

Direct Replicability

Conceptual Replicability

Scientific Inquiry

Validity

Conclusion Validity

Construct Validity

Internal Validity

External Validity

Soundness

Completeness

Degree to which results are free from errors
(Some) Process mining crimes

- Using the wrong evaluation data
  - E.g., overgeneralizing from “simplistic” logs (external validity)
- Misleading quality assessment
  - E.g., using selective measures (internal validity)
- Scientific inaccuracies
  - E.g., not evaluating all claims (construct validity)
- Improper comparison of results
  - E.g., improper treatment of competitors (conclusion validity)
- Missing information
  - E.g., only relative numbers (direct replicability)
Preventing process mining crimes

1) Be specific when reporting on your contributions.
2) Explicate assumptions.
3) Choose representative evaluation data and justify this choice.
4) Be aware of the shortcomings of quality measures.
5) Be aware of non-determinism – and don’t be afraid of statistics.
6) Make (fair) comparisons to state-of-the-art techniques.
7) Specify your computational set-up, if necessary.
8) Provide the source code and the evaluation data.

This is hard!!!
What does this mean for you?

**Read!** There are many great papers that provide (experimental) evidence for the problems I described. Make sure you are aware of the literature to avoid surprises.

**Question!** Just because someone else did an evaluation in a certain way, this doesn’t mean that you should blindly follow it.

**Justify!** You need to make many design choices when evaluating an algorithm. Be explicit about them! You should be able to explain and justify every one of them.
“If we understand better what we’re doing, we might be able to do it better”