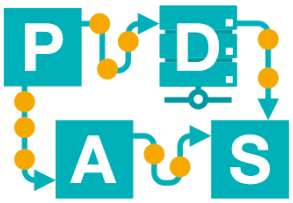


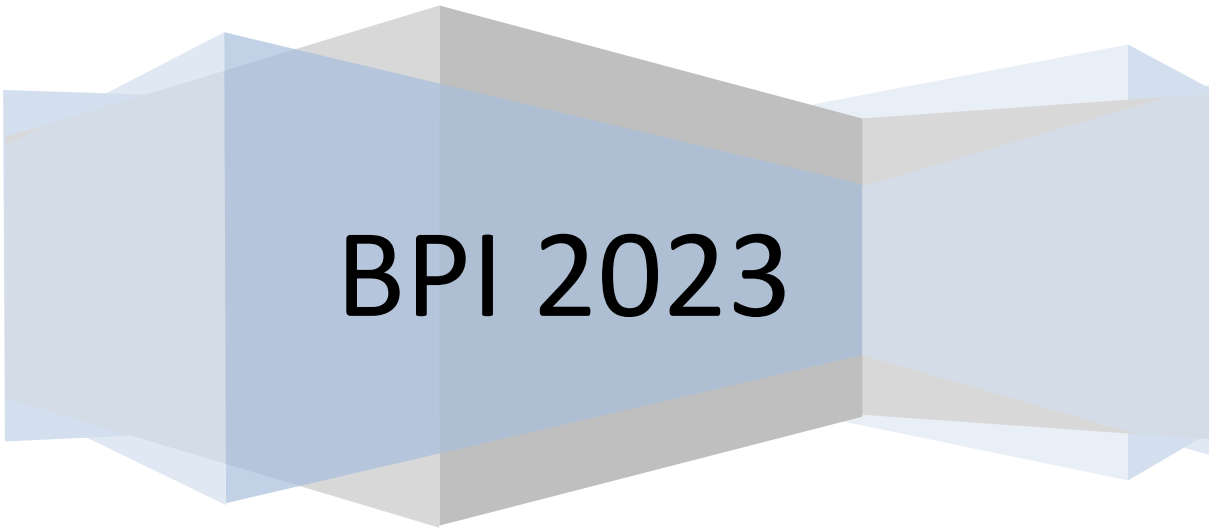
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Study Guide

Business Process Intelligence

Wil van der Aalst,
Bianka Bakullari, Harry Beyel, Nina Graves,
Benedikt Knopp, Christopher Schwanen.



BPI 2023

Table of Contents

Study Guide – Business Process Intelligence (SS 2023)

Lecturers and Instructors

Lecturer

- Prof.dr.ir. Wil van der Aalst

Instructors

- Bianka Bakullari
- Harry Beyel
- Nina Graves
- Benedikt Knopp
- Christopher Schwanen

All the questions related to the course lectures and instructions should be asked via Moodle. In case of urgent personal questions, please contact *bpi@pads.rwth-aachen.de*. Always include your matriculation number in the email to allow for fast processing. Please avoid sending e-mails to individuals or even multiple lecturers. Moreover, if you have problems with RWTH Online or RWTH Moodle that are not explicitly related to this course, please contact the persons responsible for these systems, not the lecturer or the instructors.

Course Contents and Motivation

This course starts with an overview of approaches and technologies that use *event data* to support *decision-making* and *business process (re)design*. Subsequently, the course focuses on *process mining* as a bridge between data mining and business process modeling. Business Process Intelligence (BPI) and process mining enable engineers to understand, diagnose, improve, and streamline operational processes for various organizations and systems (hospitals, banks, high-tech systems, governments, electronic shops, transportation systems, trading systems, etc.).

Process mining is part of the larger *data science* discipline. Data science aims to answer questions such as:

- What really happened? (discovery)
- Why did it happen? (root cause analysis)
- What will happen? (prediction)
- What is the best that can happen? (recommendation)

Process mining enables technology to answer such questions about operational processes in different domains. There is a huge demand for engineers having the skills and tools to turn event data into real value.

The course is at an introductory level with a practical assignment and hands-on experiences using RapidMiner, Celonis, and ProM.

The course covers the three main types of *process mining*. The first type of process mining is *process discovery*. A discovery technique takes an event log and produces a process model reflecting the behavior recorded in the log. The second type of process mining is *conformance checking*. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is *enhancement*. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event logs. While conformance checking assesses the deviances between model and reality, this third type of process mining aims at changing or extending the a-priori model. An example is enhancing a process model with performance information, e.g., showing bottlenecks. Process mining techniques can be used in offline but also online settings. The latter is known as *operational support*. An example is the detection of non-conformance at the moment the deviation actually takes

place. Another example is time prediction for running cases, i.e., given a partially executed case, the remaining processing time is estimated based on historic information of similar cases.

Process mining provides not only a bridge between data mining and business process management; it also helps to address the classical divide between "business" and "IT". *Evidence-based* business process management based on process mining helps to create a common ground for business process improvement and information systems development.

In recent years, process mining has become the primary data-driven BPM (Business Process Management) approach. Process mining is also increasingly applied in other domains (auditing, production, etc.). Data availability strengthens the necessity for appropriate data mining approaches to deal with event data. Process mining is where “Data Science” and “Process Science” meet. Currently, about 40 software vendors offer process mining tools; over half of the Fortune 500 companies and most of the larger German enterprises use process mining, and all consultancy firms provide process mining services. Many larger organizations use process mining globally (e.g., within Siemens, over 6000 people use process mining, and BMW uses process mining for 50 of its core processes). Many of the products and services you use have been "touched" by process mining software. Next to open-source tools like ProM, Pm4Py, and RapidProM there are commercial tools such as Celonis, Fluxicon, UiPath-ProcessGold, SAP-Signavio, Minit, IBM-myInvenio, QPR ProcessAnalyzer, Everflow, Software AG, Mehrwerk, Appian-LanaLabs, ABBYY Timeline, etc. The availability and application of these tools illustrate the uptake of process mining.

The course uses many examples using real-life event logs to illustrate the concepts and algorithms. After taking this course, the student can run process mining projects and understand the Business Process Intelligence (BPI) field well. Moreover, students will be able to directly apply process mining techniques in all kinds of practical settings, including internships and master projects.

Course Objectives

After taking this course, students should:

- have a good understanding of Business Process Intelligence techniques (in particular process mining),
- understand the role of Big Data and Data Science in today’s society,
- be able to relate process mining techniques to other analysis techniques such as simulation, business intelligence, data mining, machine learning, and verification,
- understand the relationship between process mining and data mining techniques like classification, clustering, and association rules,
- be able to apply basic process discovery techniques such as the alpha algorithm to learn a process model from an event log (both manually and using software tools),
- understand how more advanced process discovery techniques like region-based mining, genetic mining, and heuristic mining work,
- be able to apply basic conformance checking techniques (such as token-based replay) to compare event logs and process models (both manually and using software tools),
- be able to extend a process model with information extracted from the event log (e.g., decision points, bottlenecks),
- be able to use state-of-the-art process and data mining tools like Celonis, RapidMiner, and ProM,
- have a good understanding of the data needed to start a process mining project,
- be able to characterize the questions that can be answered based on such event data,
- explain how process mining can also be used for operational support (prediction and recommendation), and
- be able to execute process mining projects in a structured manner using the L* life-cycle model.

To show the practical relevance and to give good insights into practical challenges when dealing with real-world processes and data, there will also be a guest lecture by Angela-Sophia Gebert from Celonis on P2P and O2C processes.

Organization

The course starts on *Monday, 03.04.2022*. Lectures are planned on **Mondays from 08:30 to 10:00** and **Tuesdays from 10:30 to 12:00.** Instructions are held on **Wednesdays from 12:30 to 14:00**. There are deviations from this rule, so check the timetable at the end of this document and follow the announcements on Moodle for updates. All lectures and instructions will take place at the **AH IV (2354|030)** lecture hall.

If you face problems accessing the lectures in Moodle, you can alternatively watch the videos via

- Video AG https://video.fsmpi.rwth-aachen.de/21ss-bpi
- YouTube https://youtube.com/playlist?list=PLG_1ZxIPXO0uRZtlYxaLgc62kKfko8QQ6

Note that the content has changed significantly in comparison to the recordings above. For example, the course now uses Celonis to illustrate the different concepts and applications. This has an effect, especially on the topics related to the data perspective, such as Decision Mining, Performance and Organizational mining (not so much w.r.t. the control-flow perspective).

Lecture Material and Software

The textbook "W. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016" is the primary source of information, and the lectures will be linked to chapters in the book. It can be ordered from (Data Science in Action book) (also note the discount when registering for the Coursera or RWTH-Celonis online courses, see below). Moreover, the “Process Mining Handbook” (Process Mining Handbook link) gives a self-contained and comprehensive overview of the process mining field.

Next to the book, the following material will be provided via RWTH-Moodle:

- Lecture slides,
- Exercises,
- Event logs,
- Assignment (two parts).

The course uses the following software tools:

- **Celonis:** Obtain access to the “Celonis Academic Edition” via https://www.celonis.com/academic-signup. See https://www.celonis.com/de/academic-alliance/ for information (you must apply for an academic license).
- **ProM 6.12:** The software can be downloaded from the ProM website https://promtools.org/prom-6-12/. It requires Java 8.
- **RapidMiner 9.10:** Download the latest RapidMiner Studio version via http://rapidminer.com/educational-program/. As an RWTH student, you can apply for a license and get an unlimited version of RapidMiner (please follow the workflow on https://rapidminer.com/educational-program/ carefully and use your RWTH account).

Online Courses

As mentioned, you can watch the online lectures, e.g., via YouTube (but note that several parts have changed!): https://youtube.com/playlist?list=PLG_1ZxIPXO0uRZtlYxaLgc62kKfko8QQ6.

Study Guide – Business Process Intelligence (SS 2023)

The material of the BPI course was used for the successful MOOC (Massive Open Online Course) **Process Mining: Data Science in Action** (see https://www.coursera.org/learn/process-mining).

To learn more about the Celonis software and the connection to this course, register for the **Process Mining: From Theory to Execution** course https://www.celonis.com/wils-process-mining-class/. See also https://academy.celonis.com/.

If you register for one of the two courses mentioned, you can get a 50% discount on the e-book version of the book and a 20% discount on the hard copy.

Examination

The module comprises two parts: one **assignment** delivered in two parts (each accounting for 20%) and the **written exam** that accounts for the remaining 60% of the final grade. The assignment must be submitted in groups of 2-3 students. To pass the course, it is required to pass both the assignment and the written exam; that is, you should achieve at least 50% in the assignment and at least 50% in the final written exam.

Please note that the dates and times below are tentative and are subject to changes.

- **The written exam** (60%):
  - First option (PT1): Friday, 04/08/2023
  - Second option (PT2): Friday, 15/09/2023
- **BPI Assignment Part 1** (20%): deadline Monday, 22/05/2023 23:59 (strict!)
- **BPI Assignment Part 2** (20%): deadline Wednesday, 28/06/2023 23:59 (strict!)

Important: Only the written exam can be retaken this semester. It is not possible to retake parts of the course; therefore, the assignment results expire after the semester ends.

Plagiarism: We systematically check for plagiarism. All group members are responsible for submitting an individual piece of work and should avoid unfair academic practices. We will report cases of proven plagiarism to the examination board.

Deadlines: Deadlines are strict. Each year, a handful of students submit too late due to technical problems or have a group member uploads an incorrect or empty file. It is OK to take the risk and wait until the last moment, but do not complain afterwards. Note that working in groups is part of the learning experience. If there are internal group problems, address them internally first and do this early.

Who can take the course?

This course can be taken at the master's or bachelor's level by all students whose study program lists it as an elective course (Wahlpflichtfach).

If this does not apply to your study program, you are still welcome to participate. In this case, however, we can only provide you with a certificate showing you passed the course. If you additionally want the credits to be added to your transcript of grades, you should contact the examination board and the persons responsible for your study program. Please refrain from contacting the lecturers and instructors concerning this matter.

About the Process and Data Science (PADS) Group @ RWTH

The Process and Data Science (PADS) group, headed by prof.dr.ir. Wil van der Aalst is one of the larger research groups in the Department of Computer Science. The scope of PADS includes all activities where discrete processes are

analyzed, re-engineered, and supported in a data-driven manner. Process-centricity is combined with an array of data science techniques. The group's research and teaching activities cover the following areas: data science, process mining, business process management, data mining, process discovery, conformance checking, and simulation.

The group closely collaborates with Celonis (the leading process mining company), the Fraunhofer Institute for Applied Information Technology (FIT), larger organizations using process mining (e.g., Siemens, BMW, Philips, and Vanderlande), and consultancy firms (KPMG, Deloitte, EY, and PwC).

The primary research focuses on process mining (including process discovery, conformance checking, performance analysis, predictive analytics, operational support, and process improvement). This is combined with neighboring disciplines such as operations research, algorithms, discrete event simulation, business process management, and workflow automation.

Visit http://www.pads.rwth-aachen.de/ to learn more about our chair and possible bachelor and master theses.

Enjoy the course!

Timetable of Lectures and Instructions*

Date	Weekday	Time	#	Topic	Description
03.04.	Monday	08:30	Lecture 1	Introduction to Process Mining	Introduction to Data Science, Process Mining, and the organization of the course.
04.04.	Tuesday	10:30	Lecture 2	Decision Trees	Basic introduction to classification and decision trees.
05.04.	Wednesday	12:30	Instruction 1	Tool Introduction	ProM, RapidMiner, Celonis
11.04.	Tuesday	10:30	Lecture 3	Association Rules and Clustering	Basic introduction to unsupervised learning, frequent item sets, pattern mining, and clustering.
12.04.	Wednesday	12:30	Instruction 2	Data Mining	Lectures 2,3
17.04.	Monday	08:30	Lecture 4	Introduction to Process Discovery	A basic introduction to process discovery. What is the problem, and what approaches are possible?
18.04.	Tuesday	10:30	Lecture 5	Alpha Algorithm 1	More on Petri nets. Introduction to the Alpha algorithm.
19.04.	Wednesday	12:30	Instruction 3	Petri Nets	Lectures 4,5
24.04.	Monday	08:30	Lecture 6	Alpha Algorithm 2	Limitation and properties of the Alpha algorithm.
25.04.	Tuesday	10:30	Lecture 7	Model Quality Representation	How to evaluate discovered models? Notions like fitness, precision, generalization, and simplicity. Discussion on representations.
26.04.	Wednesday	12:30	Instruction 4	Alpha Miner	Lectures 5,6,7
02.05.	Tuesday	08:30 (AH V)	Lecture 8	Heuristic Mining	Introduction to an algorithm that can handle noise and incompleteness.
03.05.	Wednesday	12:30	Lecture 9	Region-Based Mining	Introduction to algorithms that provide guarantees but cannot handle noise.
08.05.	Monday	08:30	Lecture 10	Inductive Mining	Introduction to a state-of-the-art approach (Inductive mining).
09.05.	Tuesday	10:30	Lecture 11	Event Data and Exploration	What types of event data exist, and what are the problems when data is not "flat"?
10.05.	Wednesday	12:30	Instruction 5	Heuristic Mining and Region-Based Mining	Lectures 8,9
15.05.	Monday	08:30	Lecture 12	Conformance Checking 1	Conformance checking using footprint matrices and token-based replay
16.05.	Tuesday	10:30	Lecture 13	Conformance Checking 2	Conformance checking using token-based replay and alignments
17.05.	Wednesday	12:30	Instruction 6	Q&A: Assignment Part 1	Q&A
22.05.	Monday	08:30	Lecture 14	Decision Mining	How to learn factors that influence decisions in processes? Application of classification techniques in processes.
Monday, 22.05. 23:59: Deadline Assignment Part 1					
23.05.	Tuesday	10:30	Lecture 15	Organizational Mining and Performance	How to discover social networks from event data? How to uncover bottlenecks?
24.05.	Wednesday	12:30	Instruction 7	Inductive Mining	Lectures 10,11
05.06.	Monday	08:30	Instruction 8	Footprint and Token-Based Replay	Lecture 12
06.06.	Tuesday	10:30	Instruction 9	Alignments	Lecture 13
07.06.	Wednesday	12:30	Instruction 10	Decision Mining	Lecture 14
12.06.	Monday	08:30	Lecture 16	Refined Process Mining	Putting different things together in one framework, including prescriptive and predictive analytics. Doing a PM project. Two types of processes.
13.06.	Tuesday	10:30	Lecture 17	Process Mining Big Data	Big data, event logs and event streams, streaming process mining, decomposed process mining, process mining tools

14.06.	Wednesday	12:30	Instruction 11	Performance Analysis	Lecture 15
19.06.	Monday	08:30	Lecture 18	Guest Lecture Celonis	What are these processes all about? What are the activities? What does the data model look like? What are typical execution gaps? Etc.
20.06.	Tuesday	10:30	Lecture 19	Summary What's Next	Summary of the course. What is important for the exam? What to do next?
21.06.	Wednesday	12:30	Instruction 12	Organizational Mining	Lecture 15
26.06.	Monday	08:30	Instruction 13	Q&A: Assignment Part 2	Q&A
Wednesday, 28.06., 23:59: Deadline Assignment Part 2					
05.07.	Wednesday	12:30	Instruction 14	Celonis Case Study	Lecture 18
12.07.	Wednesday	12:30	Instruction 15	Exam Q&A	Q&A

*All lectures and instructions (except the lecture on 02.05.) are held in the **AH IV (2354|030)** lecture hall. The lecture on 02.05. will take place at **08:30** in **AH V (2356|050).**