



Chair of Process
and Data Science

RWTHAACHEN
UNIVERSITY

Study Guide

Business Process Intelligence

Wil van der Aalst, Niklas Adams, Bianka Bakullari,
Harry Beyel, Lisa Mannel, Ali Norouzifar,
Zahra Sadeghibogar, Pouya Soumand.

BPI 2022

Table of Contents

Lecturers and Instructors 2

Course Contents and Motivation 2

Course Objectives 3

Organization 4

Lecture Material and Software 4

Examination 5

Who can take the course? 5

About the Process and Data Science (PADS) Group @ RWTH..... 6

Timetable of Lectures and Instructions 7

Lecturers and Instructors

Lecturer

- Prof.dr.ir. Wil van der Aalst

Instructors

- Niklas Adams
- Bianka Bakullari
- Harry Beyel
- Lisa Mannel
- Ali Norouzifar
- Zahra Sadeghibogar
- Pouya Soumand

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. To allow for fast processing, **always include your immatriculation number in the email**. 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 specifically related to this course, please contact the persons responsible for these systems and 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 a wide variety of 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 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 is an enabling 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 also 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 the extension of a process model with performance information, e.g.,

showing bottlenecks. Process mining techniques can be used in an offline, but also online setting. 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.). The availability of data nowadays strengthens the necessity for appropriate data mining approaches that can deal with event data. Process mining is where "Data Science" and "Process Science" meet. Currently, there are about 40 software vendors offering process mining tools, over half of the Fortune 500 companies and most of the larger German enterprises use process mining, and all consultancy firms are offering process mining services. There are many larger organizations using process mining at a global scale (e.g., within Siemens over 6000 people are using process mining, and BMW is using 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 is able to run process mining projects and have a good understanding of the Business Process Intelligence (BPI) field. 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 relation 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 and Jerome Geyer-Klingenberg (TBC) from Celonis on P2P and O2C processes.

Organization

The course starts on *Monday, 04.04.2022*. Lectures are planned on **Mondays from 12:30 to 14:00** and **Tuesdays from 14:30 to 16:00**. Instructions are held on **Wednesdays from 10:30 to 12:00**. **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_1ZxIPXO0uRZtIYxaLgc62kKfko8QQ6

Note that the content has changed significantly. 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 (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" (<http://springer.com/9783662498507>) is the primary source of information and the lectures will be linked to chapters in the book. It can be ordered from <http://springer.com/9783662498507> (also note the discount when registering for the Coursera or RWTH-Celonis online courses, see below).

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 need to apply for an academic license).
- **ProM 6.10:** The software can be downloaded from the ProM web site: <http://www.promtools.org/>. It requires Java 8. For more information about the requirement of this tool, please see <https://www.promtools.org/doku.php?id=prom611>.
- **RapidMiner 9.9 (or later):** Download the latest RapidMiner Studio version via <http://rapidminer.com/educational-program/>. Note that as a 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 last year's lecture online, e.g., via YouTube (but note that several parts have changed!): https://youtube.com/playlist?list=PLG_1ZxIPXO0uRZtIYxaLgc62kKfko8QQ6.

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 of the book.

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): **29/07/2022** (tentative, check later)
 - Second option (PT2): **09/09/2022** (tentative, check later)
- **BPI Assignment Part 1** (20%): deadline **25/05/2022 23:59 (strict!)**
- **BPI Assignment Part 2** (20%): deadline **04/07/2022 23:59 (strict!)**

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

Plagiarism: We systematically check for plagiarism. All group members are responsible to submit 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 there are again a handful of students that submit too late due to technical problems or that have a group member that uploads an incorrect or empty file. It is OK if you take the risk and wait until the last moment, but do not complain afterward. Note that working in groups is part of the learning experience. If there are internal group problems, address them yourself and do this early.

Who can take the course?

This course can be taken at the master or bachelor level from 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 that you passed the course. **If you additionally want the credits to be added to your transcript of grades, you should get in contact with the examination board and the persons responsible for your study program. Please refrain from contacting the lecturers and instructors with respect to 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).

Currently, the main research focus is 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 find out about possible bachelor and master theses.

Enjoy the course!

Timetable of Lectures and Instructions

#	Topic	Description	Date	book chapters	mooc lectures
Lecture 1	Introduction to Process Mining	Introduction to Data Science, Process Mining, and the organization of the course.	Monday, 04.04.2022	1 & 2	1.1-1.3
Lecture 2	Data Science Basics (1/2)	Basic introduction to classification and decision trees (intended for those not having a data mining background).	Tuesday, 05.04.2022	4	1.4-1.5
Instruction 1	Tool Introduction	ProM and RapidMiner	Wednesday, 06.04.2022	-	-
Lecture 3	Data Science Basics (2/2)	Basic introduction to unsupervised learning, frequent item sets, pattern mining, and clustering (for those not having a data mining background).	Monday, 11.04.2022	4	1.6-1.8
Lecture 4	Introduction to Process Discovery	Basic introduction to process discovery. What is the problem and what approaches are possible?	Tuesday, 12.04.2022	2 & 3	1.2-2.1
Lecture 5	Petri Nets & Alpha Algorithm	More on Petri nets. Introduction to the Alpha algorithm.	Wednesday, 13.04.2022	3, 5 & 6	2.2-2.7
Instruction 2	Data Mining + RapidMiner	Lectures 2,3	Tuesday, 19.04.2022	-	-
Instruction 3	Petri Nets	Lectures 4,5	Wednesday, 20.04.2022	-	-
Lecture 6	Alpha Algorithm Continued	Limitation and properties of the Alpha algorithm.	Monday, 25.04.2022	6	2.5-2.7
Lecture 7	Quality of Discovered Models and Representations	How to evaluate discovered models? Notions like fitness, precision, generalization, simplicity. Discussion on representations.	Tuesday, 26.04.2022	6	3.1-3.2
Instruction 4	Alpha Algorithm and Model Evaluation	Lectures 5,6,7	Wednesday, 27.04.2022	-	-
Lecture 8	Heuristic Mining	Introduction to an algorithm that can handle noise and incompleteness.	Monday, 02.05.2022	7	
Lecture 9	Region-Based Mining	Introduction to algorithms that provide guarantees but cannot handle noise.	Tuesday, 03.05.2022	7	3.7-4.2
Instruction 5	Heuristic Mining and Region-Based Mining	Lectures 8,9	Wednesday, 04.05.2022	-	-
Lecture 10	Inductive Mining	Introduction to a state-of-the-art approach (Inductive mining).	Monday, 09.05.2022	7	not in MOOC
Lecture 11	Event Data and Exploration	What types of event data exist and what are the problems when data is "not flat"?	Tuesday, 10.05.2022	5	4.8, 6.2
Instruction 6	Advanced Process Discovery Techniques	Lectures 10,11	Wednesday, 11.05.2022	-	-
Lecture 12	Conformance Checking (1/2)	Conformance checking using footprint matrices and token based replay	Monday, 16.05.2022	8	4.4-4.6
Lecture 13	Conformance Checking (2/2)	Conformance checking using token based replay and alignments	Tuesday, 17.05.2022	8	4.5-4.7
Instruction 7	Q&A Assignment Part 1	Assignment Part 1	Wednesday, 18.05.2022	-	-
Instruction 8	Conformance Checking	Lecture 12	Tuesday, 24.05.2022	-	-
Instruction 9	Conformance Checking	Lecture 13	Wednesday, 25.05.2022	-	-
25.05.2022, 23:59: Deadline Assignment Part 1					
Lecture 14	Decision Mining	How to learn factors influencing decisions in processes? Application of classification techniques in processes.	Monday, 30.05.2022	9	5.2-5.3
Lecture 15	Organizational Mining & Bottleneck Analysis	How to discover social networks from event data? How to uncover bottlenecks?	Tuesday, 31.05.2022	11,12	not in MOOC
Instruction 10	Decision Mining	Lecture 14	Wednesday, 01.06.2022	-	-
Lecture 16	Guest Lecture Celonis: Improving O2C and P2P supported by ERP systems like SAP in Practice	What are these processes all about? What are the activities? What does the data model look like? What are typical execution gaps? Etc.	Monday, 13.06.2022	-	not in MOOC

Study Guide – Business Process Intelligence (SS 2022)

Lecture 17	Refined Process Mining Framework and Operational Support	Putting different things together in one framework, including prescriptive and predictive analytics. Doing a PM project. Two types of processes.	Tuesday, 14.06.2022	10	5.9-6.1
Instruction 11	Organizational Mining	Lecture 15	Wednesday, 15.06.2022	-	-
Lecture 18	Dealing with Big Event Data	Big data, event logs and event streams, streaming process mining, decomposed process mining, process mining tools	Monday, 20.06.2022	11,12	not in MOOC
Lecture 19	Discussion of an old/possible exam	Information will be provided before.	Tuesday, 21.06.2022	13,14	6.4-6.7
Instruction 12	Performance and Bottleneck Analysis	Lecture 15	Wednesday, 22.06.2022	-	-
Lecture 20	Summary of the Course and Next Steps	Summary of the course. What is important for the exam? What to do next?	Monday, 27.06.2022	15,16	6.8-6.9
Instruction 13	Refined Process Mining	Lecture 17	Tuesday, 28.06.2022	-	-
Instruction 14	Q&A Assignment Part 2	Assignment Part 2	Wednesday, 29.06.2022	-	-
04.07.2022, 23:59: Deadline Assignment Part 2					
Instruction 15	Q&A	Exam Q&A	Wednesday, 13.07.2022	15,16	6.8-6.9