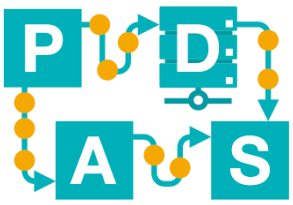


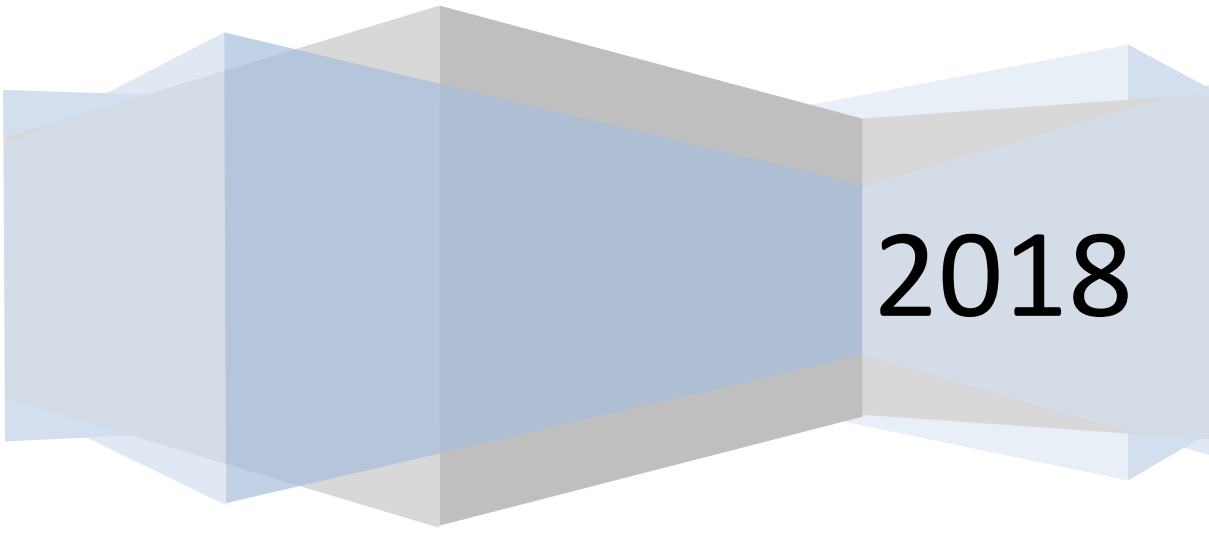
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Study Guide Business Process Intelligence

Wil van der Aalst

Mustafa Ghani & Mohammadreza Fani Sani



2018

Study Guide Business Process Intelligence (SS 2018)

Lecturers

- prof.dr.ir. Wil van der Aalst (lectures)
- Mustafa Ghani (instructions & assignment)
- Mohammadreza Fani Sani (instructions & assignment)

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 various practical assignments.

The course covers the three main types of *process mining*. The first type of process mining is *discovery*. A discovery technique takes an event log and produces a process model without using any a-priori information. An example is the alpha-algorithm that takes an event log and produces a Petri net explaining the behavior recorded in the log. The second type of process mining is *conformance*. 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 log. Whereas conformance checking measures the alignment 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 attention for Big Data and

the uptake of data science strengthen this development. Process mining is where “Data Science” and “Process Science” meet! Currently, there are about 25 software vendors offering process mining tools. Next, to Disco (Fluxicon’s tool is used in the course next to the open-source tool ProM), tools like Celonis Process Mining, ProcessGold Enterprise Platform, Minit, myInvenio, QPR ProcessAnalyzer, Lexmark’s Perceptive Process Mining and many more are now available. 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, one 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.

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 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 tools),
- be able to extend a process model with information extracted from the event log (e.g., show bottlenecks),
- 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.

Organization & Lecture Material

The course starts on April 11th 2018. Lectures are on Tuesdays and Wednesdays from 8.30 to 10.00 in 2350|314.1 (AH III). Instructions are on Fridays from 8.30 to 10.00 in 2350|314.1 (AH III).
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. Make sure to order it in time, e.g., via http://springer.com/9783662498507, http://amzn.com/3662498502, or https://www.bol.com/nl/p/process-mining/9200000057066388.

Next to the book, the following material will be distributed via the RWTH E-learning platform and www.processmining.org:

- Slides,
- Exercises,
- Event logs, and
- Assignments.

Video Lectures

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). Over 100.000 participants joined the course over the last two years, illustrating the global interest in the topic**.** This way you can watch lectures and do additional assignments at your own pace (only as background or in case things are not clear).

- Lecture 1.1: Data Science and Big Data (17 min.)
- Lecture 1.2: Different Types of Process Mining (21 min.)
- Lecture 1.3: How Process Mining Relates to Data Mining (20 min.)
- Lecture 1.4: Learning Decision Trees (27 min.)
- Lecture 1.5: Applying Decision Trees (21 min.)
- Lecture 1.6: Association Rule Learning (18 min.)
- Lecture 1.7: Cluster Analysis (13 min.)
- Lecture 1.8: Evaluating Mining Results (15 min.)

- Lecture 2.1: Event Logs and Process Models (14 min.)
- Lecture 2.2: Petri Nets (1/2) (16 min.)
- Lecture 2.3: Petri Nets (2/2) (18 min.)
- Lecture 2.4: Transition Systems and Petri Net Properties (21 min.)
- Lecture 2.5: Workflow Nets and Soundness (17 min.)
- Lecture 2.6: Alpha Algorithm: A Process Discovery Algorithm (25 min.)
- Lecture 2.7: Alpha Algorithm: Limitations (23 min.)
- Lecture 2.8: Introducing ProM and Disco (25 min.)

- Lecture 3.1: Four Quality Criteria For Process Discovery (19 min.)
- Lecture 3.2: On The Representational Bias of Process Mining (17 min.)
- Lecture 3.3: Business Process Model and Notation (BPMN) (15 min.)
- Lecture 3.4: Dependency Graphs and Causal Nets (21 min.)
- Lecture 3.5: Learning Dependency Graphs (21 min.)
- Lecture 3.6: Learning Causal nets and Annotating Them (18 min.)
- Lecture 3.7: Learning Transition Systems (15 min.)
- Lecture 3.8: Using Regions to Discover Concurrency (18 min.)

- Lecture 4.1: Two-Phase Process Discovery And Its Limitations (15 min.)
- Lecture 4.2: Alternative Process Discovery Techniques (23 min.)
- Lecture 4.3: Introduction to Conformance Checking (12 min.)
- Lecture 4.4: Conformance Checking Using Causal Footprints (10 min.)
- Lecture 4.5: Conformance Checking Using Token-Based Replay (15 min.)
- Lecture 4.6: Token-Based Replay: Some Examples (15 min.)
- Lecture 4.7: Aligning Observed and Modeled Behavior (18 min.)
- Lecture 4.8: Exploring Event Data (21 min.)

- Lecture 5.1: About the Last Two Weeks of This Course (10 min.)
- Lecture 5.2: Mining Decision Points (17 min.)
- Lecture 5.3: Discovering Data Aware Petri Nets (12 min.)
- Lecture 5.4: Mining Bottlenecks (11 min.)
- Lecture 5.5: Mining Social Networks (17 min.)
- Lecture 5.6: Organizational Mining (9 min.)
- Lecture 5.7: Combining Different Perspectives (13 min.)
- Lecture 5.8: Comparative Process Mining Using Process Cubes (13 min.)
- Lecture 5.9: Refined Process Mining Framework (11 min.)

- Lecture 6.1: Operational Support: Detect, Predict and Recommend (17 min.)
- Lecture 6.2: Getting the Right Event Data (17 min.)
- Lecture 6.3: Guidelines for Logging (10 min.)
- Lecture 6.4: Process Mining Software (16 min.)
- Lecture 6.5: How to Conduct a Process Mining Project (11 min.)
- Lecture 6.6: Mining Lasagna Processes (6 min.)
- Lecture 6.7: Mining Spaghetti Processes (8 min.)
- Lecture 6.8: Process Models as Maps (12 min.)
- Lecture 6.9: Data Science in Action (9 min.)

Software

The course uses the following analysis tools:

- **ProM lite 1.2:** The software can be downloaded from the ProM web site: http://www.promtools.org/ (see http://www.promtools.org/doku.php?id=promlite112). It requires Java 7.0 or higher. For more information about requirement of this tool, please see https://svn.win.tue.nl/trac/prom/wiki/ProMLite12/ReleaseNotes.
- **Disco 2.1.0:** Download from http://fluxicon.com/academic/, and visit http://fluxicon.com/academic/material/ (use RWTH email address!)
- **Celonis:** Obtain access via https://academiccloud.celonis.com. See https://www.celonis.com/de/academic-alliance/ for information (you need to apply for an academic license).
- **RapidMiner 8.0:** 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).

Also, note that the RapidProM extension www.rapidprom.org is not needed for this course, but will be used in later specialized courses and may already be interesting.

Examination

Next to the final written test, there are two assignments each counting for 20% of the final result. The final test counts for remaining 60% of the final result:

- **Final written test** (60%):
  - First option (PT1): Friday 27/07/2018 (9.00-12.00) 2350|111 (AH II)
  - Second option (PT2): Friday 14/09/2018 (9.00-12.00) 2350|111 (AH II)
- **Assignment 1** (20%): deadline Wednesday 20/05/2018
- **Assignment 2** (20%): deadline Thursday 15/07/2018

Participation in the assignments is required for participation in the final test. Only the final test can be retaken in this semester if the first option is chosen. Assignments can only be redone in the next academic year.

Detailed descriptions of the assignments will be handed out separately.

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 research units in the Department of Computer Science. The scope of PADS includes all activities where discrete processes are analyzed, reengineered, and/or supported in a data-driven manner. Process-centricity is combined with an array of Data Science techniques. The group's research and teaching activities can be characterized by the keywords: Data Science, Process Science, Process Mining, Business Process Management, Data Mining, Process Discovery, Conformance Checking, and Simulation.

The group has been established in the context of the Alexander von Humboldt Professorship awarded to prof.dr.ir Wil van der Aalst in 2017. The award is Germany's most prestigious and valuable prize for international researchers. The PADS group supports RWTH's strategy to further strengthen its Data Science capabilities. The group also closely collaborates with the Fraunhofer Institute for Applied Information Technology (FIT).

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 possible Bachelor and master theses.

Planning

See spreadsheet.

#	Lecture	date	day	description	book chapters	mooc lectures
1	Introduction to Process Mining	11.04.2018	Wednesday	Introduction to Data Science, Process Mining, and the organization of the course.	1 & 2	1.1-1.3
2	Decision Trees	13.04.2018	Friday	Basic introduction to classification and decision trees (intended for those not having a data mining background).	4	1.4-1.5
3	Association Rules & Clustering	17.04.2018	Tuesday	Basic introduction to unsupervised learning, frequent item sets, pattern mining, and clustering (for those not having a data mining background).	4	1.6-1.8
Instruction 1	Decision Tree & Entropy	18.04.2018	Wednesday	Introduction to instruction lectures & some exercises about Entropy, Information gain, Decision tree and RapidMiner.		
Instruction 2	Association Rules & Clustering	20.04.2018	Friday	Some exercises about Association rules and clustering and practical experiment with RapidMiner.		
4	Introduction to Process Discovery	24.04.2018	Tuesday	Basic introduction to process discovery. What is the problem and what approaches are possible?	2 & 3	1.2-2.1
5	Petri Nets & Alpha Algorithm	25.04.2018	Wednesday	More on Petri nets. Introduction to the Alpha algorithm.	3, 5 & 6	2.2-2.7
Instruction 3	Petri net exercises	27.04.2018	Friday	Examples about Petri nets and features like soundness and also getting familiar with Disco.		
6	Alpha Algorithm Continued	02.05.2018	Wednesday	Limitation and properties of the Alpha algorithm.	6	2.5-2.7
Instruction 4	Alpha Algorithm	04.05.2018	Friday	Applying the Alpha mining algorithm on some example event logs and also get familiar with Prom Lite.		
Instruction 5	Casual NET	11.05.2018	Friday	Discovering some casual net from example event logs.		
7	Quality of Discovered Models and Representations	15.05.2018	Tuesday	How to evaluate discovered models? Notions like fitness, precision, generalization, simplicity. Discussion on representations.	6	3.1-3.2
8	Heuristic Mining	16.05.2018	Wednesday	Introduction to an algorithm that can handle noise and incompleteness.	7	3.3-3.6
Instruction 6	Transition System and Q&A Session for Assignment 1	18.05.2018	Friday	Answering questions of students related to Assignment 1 and providing some exercises related to transition systems and Petri nets.		
9	Region-Based Mining	29.05.2018	Tuesday	Introduction to algorithms that provide guarantees but cannot handle noise.	7	3.7-4.2
Instruction 7	Detecting Regions	30.05.2018	Wednesday	Some examples to learn the region based mining algorithm.		
10	Inductive Mining	01.06.2018	Friday	Introduction to a state-of-the-art approach (Inductive mining).	7	not in MOOC
11	Event Data and Exploration	06.06.2018	Wednesday	What types of event data exist and what are the problems when data is "not flat"?	5	4.8 6.2
Instruction 8	Inductive Mining Examples	08.06.2018	Friday	Some examples to learn the inductive mining algorithm.		
12	Conformance Checking (1/2)	13.06.2018	Wednesday	Conformance checking using footprint matrices and token based replay.	8	4.4-4.6
Instruction 9	Token-Based Replay	15.06.2018	Friday	Exercises about move on log, move on model and replay.		
13	Conformance Checking (2/2)	19.06.2018	Tuesday	Conformance checking using token based replay and alignments.	8	4.5-4.7
14	Decision Mining	20.06.2018	Wednesday	How to learn factors influencing decisions in processes? Application of classification techniques in processes.	9	5.2-5.3
Instruction 10	Conformance Checking Exercises	22.06.2018	Friday	Exercises to compute fitness based on replay.		
Instruction 11	Social Network Analysis & Resource Activity Matrix	29.06.2018	Friday	Some exercise for discovering different types of social networks and analyzing them.		
15	Organizational Mining & Bottleneck Analysis	03.07.2018	Tuesday	How to learn about bottlenecks and the distribution of work in processes?	9	5.4-5.8
16	Refined Process Mining Framework and Operational Support	04.07.2018	Wednesday	Putting different things together in one framework, including prescriptive and predictive analytics.	10	5.9-6.1
Instruction 12	Performance Analysis	06.07.2018	Friday	Finding bottlenecks of the process model.		
17	Process Mining Software & Big Data	10.07.2018	Tuesday	How to deal with the four V's of Big data? How to decompose process mining problems?	11,12	not in MOOC
Instruction 13	Question & Answer Session for Assignment 2	13.07.2018	Friday	Answering to question of students about Assignment 2 and other process mining tools.		
18	Process Mining Projects and Different Types of Processes	17.07.2018	Tuesday	A more practical perspective on process mining. How to approach a project? What types of processes exist?	13,14	6.4-6.7
19	Summary and Old Exam	18.07.2018	Wednesday	Summary of the course. What is important for the exam? What to do next?	15,16	6.8-6.9
Instruction 14	Answering Questions about the exam	20.07.2018	Friday	Answering remaining questions related to the whole course and discussion of old/possible exam questions.		

	Announcement of Assignment 1	24.04.2018
	Deadline for Assignment 1	20.05.2018
	Announcement of Assignment 2	01.06.2018
	Deadline for Assignment 2	15.07.2018