Starts on November 12th 2014,   **register via https://www.coursera.org/course/procmin**

Process Mining: Data science in Action

Process mining is the missing link between model-based process analysis and data-oriented analysis techniques. Through concrete data sets and easy to use software the course provides data science knowledge that can be applied directly to analyze and improve processes in a variety of domains.

Take the first online process mining course, for free!

About the Course

Data science is the profession of the future, because organizations that are unable to use (big) data in a smart way will not survive. It is not sufficient to focus on data storage and data analysis. The data scientist also needs to relate data to process analysis. **Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining.** Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand-made or discovered automatically). This technology has become available only recently, but it can be applied to any type of operational processes (organizations and systems). Example applications include: analyzing treatment processes in hospitals, improving customer service processes in a multinational, understanding the browsing behavior of customers using a booking site, analyzing failures of a baggage handling system, and improving the user interface of an X-ray machine. All of these applications have in common that dynamic behavior needs to be related to process models. Hence, we refer to this as "data science in action".

The course explains the key analysis techniques in process mining. Participants will learn various process discovery algorithms. These can be used to automatically learn process models from raw event data. Various other process analysis techniques that use event data will be presented. Moreover, the course will provide **easy-to-use software**, **real-life data sets**, and **practical skills** to **directly apply the theory** in a variety of application domains.

Course Syllabus

This course starts with an overview of approaches and technologies that use event data to support decision making and business process (re)design. Then the course focuses on **process mining as a bridge between data mining and business process modeling**. The course is at an introductory level with various practical assignments.

The course covers the three main types of process mining.

1. 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 process model (a Petri net) explaining the behavior recorded in the log.
2. 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.
3. 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.

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 field.

After taking this course you should:
- have a good understanding of **Business Process Intelligence** techniques (in particular process mining),
- understand the role of **Big Data** 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,
- be able to apply basic process discovery techniques to learn a process model from an event log (both manually and using tools),
- be able to apply basic **conformance checking** techniques 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 **conduct process mining projects** in a structured manner.

Recommended Background

A basic understanding of logic, sets, and statistics (at the undergraduate level) is assumed. Basic computer skills are required to use the software provided with the course (but no programming experience is needed). Participants are also expected to have an interest in process modeling and data mining but no specific prior knowledge is assumed as these concepts are introduced in the course.

Suggested Readings

No required texts. Although the lectures are designed to be self-contained, we recommend (but do not require) that students use the book "Process Mining: Discovery, Conformance and Enhancement of Business Processes by W.M.P. van der Aalst, Springer Verlag, 2011 (ISBN 978-3-642-19344-6)", which is closely aligned with this course. One can read the process mining manifesto or visit the web site http://www.processmining.org/ to see more background material.

Course Format

- Course starts on November 12th 2014, register via https://www.coursera.org/course/procmin.
- The course consists of **6 weeks**. Every week consists of a series of short lecture videos (called modules) of about 8-15 minutes each.
- Reading assignments are provided for every week.
- Forum discussions.
- Main assignments where concrete data sets are analyzed using process mining software.
- Final exam (multiple choice, online).

FAQ

· **Data Scientist: The Sexiest Job of the 21st Century?**
Hal Varian, the chief economist at Google said in 2009:"The sexy job in the next 10 years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?". Later the article "Data Scientist: The Sexiest Job of the 21st Century" triggered a discussion on the emerging need for data scientists. This was picked up by several media and when analyzing job vacancies, one can indeed see the rapidly growing demand for data scientists. The recent attention for Big Data illustrates the importance of data science.

· **Is process mining the same as data mining?**
Traditional data mining approaches are not process-centric. Input for data mining is typically a set of records and the output is a decision tree, a collection of clusters, or frequent patterns. Process mining starts from events and the output is related to an end-to-end process model. Data mining tools can be used to support particular decisions in a larger process. However, they cannot be used for process discovery, conformance checking, and other forms of process analysis. The course also introduces basic data mining approaches and relates these to process mining to show differences and commonalities.

· **What kind of software will be used?**
The courses uses *ProM*, an open-source process mining framework (see www.processmining.org), and *Disco*, a commercial process mining tool from Fluxicon (see www.fluxicon.com). Disco is an easy to use tool that can be used free of charge by course participants. Using Disco it is very easy to convert raw data into an event log suitable for process mining and quickly create process models showing the bottlenecks in a process. ProM is a more advanced tool that provides hundreds of different types of analysis. All process mining techniques discussed in the course are supported by ProM.

· **What kind of datasets will be used?**
The course will provide several data sets ranging from simple synthetic event logs to complex and large real-life event logs, e.g., treatment data of a hospital, incident logs from a car manufacturer, loan application logs from an insurance company, and event logs from a bank. The simple event logs are used to explain and illustrate the techniques. The complex event logs are used to provide insights into the challenges real-life data science projects are facing.

· **Is process mining only suitable for the analysis of business processes?**
No, although many of the examples will come from business processes, one can also find processes in software and all kinds of devices. Process mining can for example also be used to understand why and when machines and software products fail. Through the internet of things more and more devices will be connected to the internet, thus significantly extending the reach of process mining. Process mining can be used for the analysis of any behavior, i.e., also at the level of machines and hardware/software systems.

· **Can I apply process mining to my own data?**
Event data is everywhere, as is illustrated by the many examples in this course. Participants are encouraged to apply the software to data sets surrounding them, e.g., data taken from social media (twitter and facebook) or from enterprise information systems surrounding them (e.g., SAP).

Visit

https://www.coursera.org/course/procmin

for more information