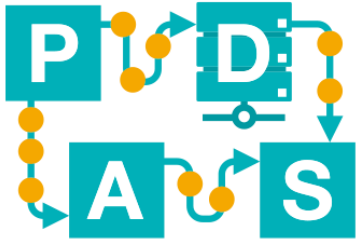
Decision Mining

Harry Beyel

BPI-I10



Chair of Process and Data Science

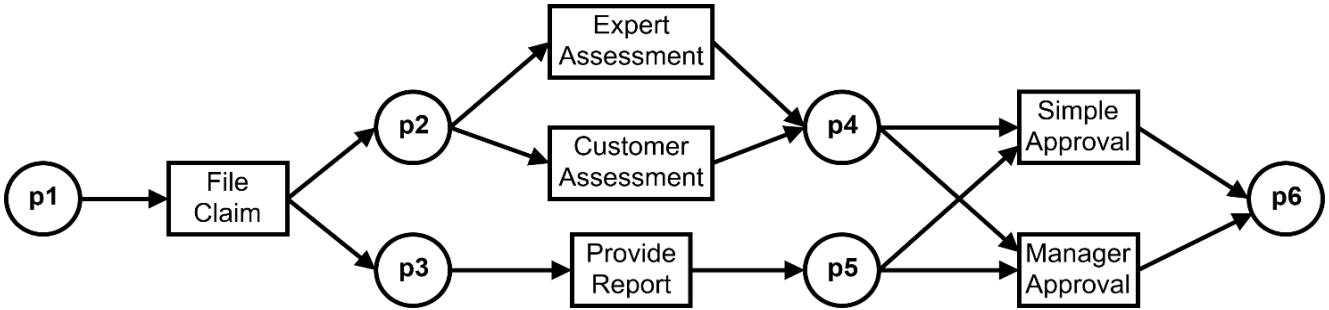
RWTH AACHEN UNIVERSITY

Goals

1. Given a process model, identify decision points.
2. Understand the difference between case-level and event-level decision points.
3. For each decision point, identify all possible choices.
4. Create a decision tree using decision outcomes as the response variable. Add the guards to the model.
5. Given the decision tree and the values of predictor variables of a new instance, predict the outcome of the decision.

Exercise 1

a) The following process model shows the insurance claim handling process in case of water damage. Identify the decision points and their possible outcomes.



Exercise 1

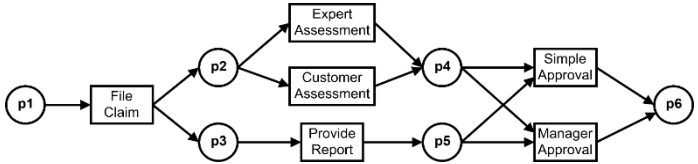
b) Given the corresponding event log, determine which attributes are case attributes.

Case ID	Activity	Transition	Timestamp	Coverage	Cost	Type
1	File Claim	complete	15.12.2020	Extended	20000	Black
1	Expert Assessment	complete	19.12.2020	Extended	20000	Black
1	Provide Report	complete	20.12.2020	Extended	20000	Black
1	Manager Approval	complete	22.01.2021	Extended	20000	Black
2	File Claim	complete	05.01.2021	Standard	2400	Black
2	Customer Assessment	complete	06.01.2021	Standard	2400	Black
2	Provide Report	complete	10.01.2021	Standard	2400	Black
2	Simple Approval	complete	21.01.2021	Standard	2400	Black
3	File Claim	complete	06.01.2021	Extended	35000	Black
3	Provide Report	complete	08.01.2021	Extended	35000	Black
3	Expert Assessment	complete	09.01.2021	Extended	35000	Black
3	Manager Approval	complete	25.01.2021	Extended	35000	Black
4	File Claim	complete	07.01.2021	Standard	1900	Black
4	Provide Report	complete	08.01.2021	Standard	1900	Black
4	Customer Assessment	complete	12.01.2021	Standard	1900	Black
4	Simple Approval	complete	18.01.2021	Standard	1900	Black
5	File Claim	complete	12.01.2021	Extended	12000	Clean
5	Provide Report	complete	16.01.2021	Extended	12000	Clean
5	Expert Assessment	complete	17.01.2021	Extended	12000	Clean
5	Simple Approval	complete	28.01.2021	Extended	12000	Clean
6	File Claim	complete	14.01.2021	Standard	15000	Black
6	Provide Report	complete	15.01.2021	Standard	15000	Black
6	Customer Assessment	complete	16.01.2021	Standard	15000	Black
6	Manager Approval	complete	02.02.2021	Standard	15000	Black

Exercise 1

c) Consider the type of approval as the outcome of each process run. Suppose you want to create a classification problem where the type of approval is the target variable.
What are the instances, and what could be predictor variables?

Case ID	Activity	Transition	Timestamp	Coverage	Cost	Type
1	File Claim	complete	15.12.2020	Extended	20000	Black
1	Expert Assessment	complete	19.12.2020	Extended	20000	Black
1	Provide Report	complete	20.12.2020	Extended	20000	Black
1	Manager Approval	complete	22.01.2021	Extended	20000	Black
2	File Claim	complete	05.01.2021	Standard	2400	Black
2	Customer Assessment	complete	06.01.2021	Standard	2400	Black
2	Provide Report	complete	10.01.2021	Standard	2400	Black
2	Simple Approval	complete	21.01.2021	Standard	2400	Black
3	File Claim	complete	06.01.2021	Extended	35000	Black
3	Provide Report	complete	08.01.2021	Extended	35000	Black
3	Expert Assessment	complete	09.01.2021	Extended	35000	Black
3	Manager Approval	complete	25.01.2021	Extended	35000	Black
4	File Claim	complete	07.01.2021	Standard	1900	Black
4	Provide Report	complete	08.01.2021	Standard	1900	Black
4	Customer Assessment	complete	12.01.2021	Standard	1900	Black
4	Simple Approval	complete	18.01.2021	Standard	1900	Black
5	File Claim	complete	12.01.2021	Extended	12000	Clean
5	Provide Report	complete	16.01.2021	Extended	12000	Clean
5	Expert Assessment	complete	17.01.2021	Extended	12000	Clean
5	Simple Approval	complete	28.01.2021	Extended	12000	Clean
6	File Claim	complete	14.01.2021	Standard	15000	Black
6	Provide Report	complete	15.01.2021	Standard	15000	Black
6	Customer Assessment	complete	16.01.2021	Standard	15000	Black
6	Manager Approval	complete	02.02.2021	Standard	15000	Black



Exercise 1

d) With the chosen predictor variables, create a table such that each row represents an instance and there is a column for the outcome (type of approval) and a column for each predictor variable.

Exercise 1

e) Create a decision tree that explains the type of approval in terms of predictor variables.

Exercise 1

f) Assume that damage costs are grouped into *High* (>10000) and *Low* (≤ 10000). Repeat tasks d) and e) by mapping each cost value onto *High* or *Low*.

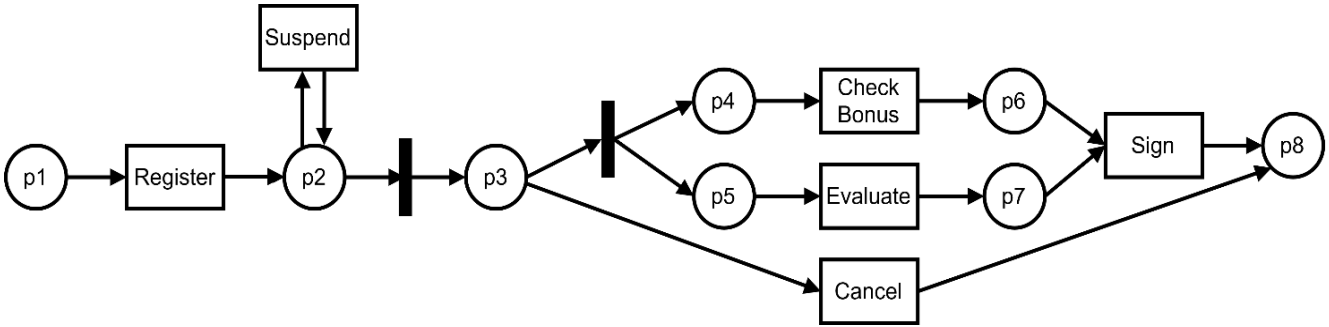
Exercise 1

Note: There may be more than one decision tree that correctly classifies all instances. The correct solution is the one where the predictor variables are added in the order that *maximizes information gain*.

E.g., try constructing a decision tree for f) where the root attribute is *Coverage* and another one where the root attribute is *Type*.

Exercise 2

a) The given model shows the process of changing the electricity supplier. Each case corresponds to a customer that decides to change their electricity supplier. Identify the decision points and their possible outcomes.



Exercise 2

The event data to the right shows the corresponding event log.

Case ID	Activity	Transition	Timestamp	Resource
1	Register	complete	08.01.2021	Mike
1	Check Bonus	complete	09.01.2021	John
1	Evaluate	complete	11.01.2021	Maria
1	Sign	complete	14.01.2021	Maria
2	Register	complete	09.01.2021	James
2	Suspend	complete	10.01.2021	Klaus
2	Suspend	complete	12.01.2021	Klaus
2	Cancel	complete	14.01.2021	John
3	Register	complete	11.01.2021	James
3	Suspend	complete	12.01.2021	John
3	Check Bonus	complete	13.01.2021	Maria
3	Evaluate	complete	14.01.2021	Maria
3	Sign	complete	15.01.2021	John
4	Register	complete	17.01.2021	Mike
4	Suspend	complete	18.01.2021	John
4	Suspend	complete	20.01.2021	John
4	Evaluate	complete	22.01.2021	Klaus
4	Check Bonus	complete	23.01.2021	John
4	Sign	complete	25.01.2021	Klaus
5	Register	complete	19.01.2021	James
5	Suspend	complete	20.01.2021	Maria
5	Suspend	complete	21.01.2021	Maria
5	Suspend	complete	22.01.2021	Klaus
5	Cancel	complete	24.01.2021	John
6	Register	complete	21.01.2021	James
6	Evaluate	complete	22.01.2021	John
6	Check Bonus	complete	23.01.2021	Klaus
6	Sign	complete	26.01.2021	Klaus

Exercise 2

b) We want to determine how the resource who executed the activity “Register” and the number of executions of activity “Suspend” affect whether the change is cancelled (activity “Cancel”) or not.
Create a table containing the case ID, the name of the resource who executed “Register”, the number of executions of “Suspend”, and the outcome (Cancel or no Cancel).

Exercise 2

c) Using the information from the table created in b), discover a decision tree that correctly classifies all instances w.r.t. the outcome (Cancel / no Cancel).
Add the discovered guards to the model.

Decision Mining with Silent Transitions

Process model of event log without silent transitions:

- → For each decision point, there is an event in the log whose activity indicates the choice at that decision point.
- → If the decision point is not in a loop:
  - → There is precisely one choice for each case. One can create a case-level classification problem (a case-based situation table) and use case-level variables as predictor variables.
  - → One can also create an event-level classification problem (an event-based situation table) and use the variables related to the events indicating the choice as predictor variables.
- → If the decision point is in a loop:
  - → There may be multiple decisions within the same case. One has to create an event-level situation table and use the variables related to the events indicating choices as predictor variables.

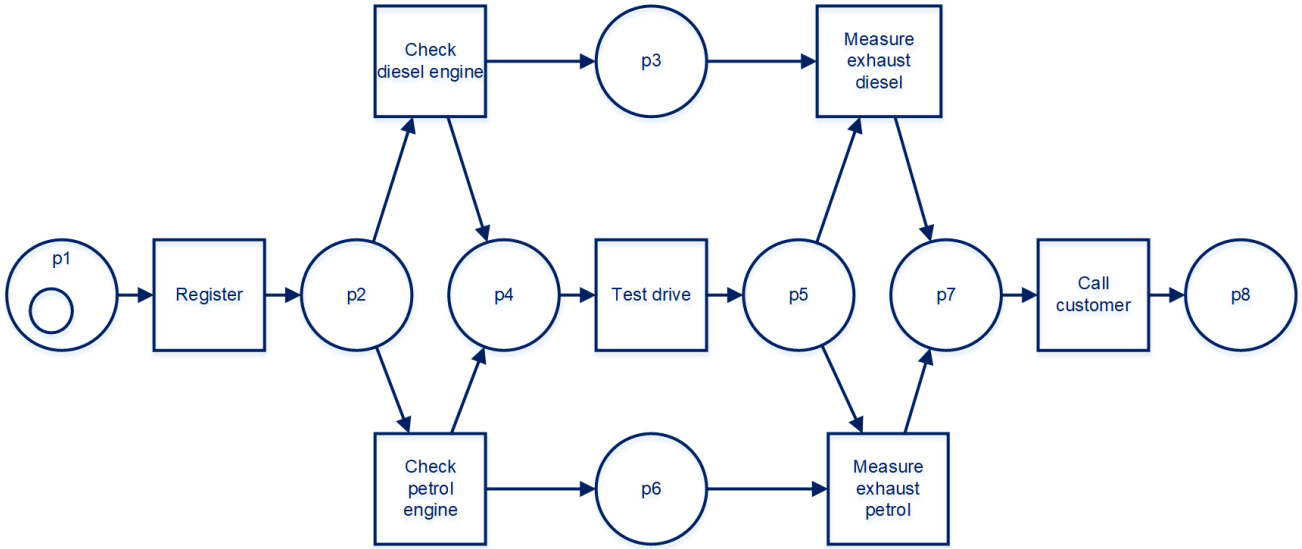
Decision Mining with Silent Transitions

Process model of event log with silent transitions:

- → **If a silent transition is one of the choices at a decision point, there is no specific event in the data that indicates this “silent” transition.**
- → **You can notice the choice of the silent transition whenever the other “visible” choices are missing in a trace.**
- → **Another way: Determine the first visible transitions that are enabled if and only if the silent transition in the choice fires. If one of these visible transitions appears in the log, their occurrence indicates the choice of the silent transition.**

Exercise 3

a) Where are the decision points in the given process model?

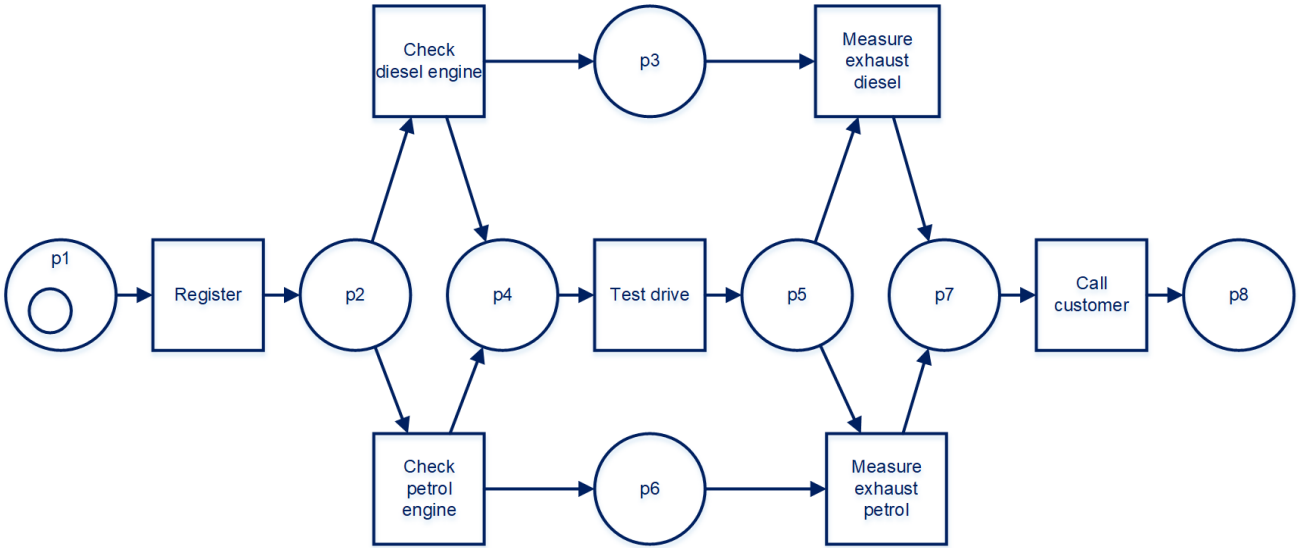


Exercise 3

b) Suppose a case in the process has the following trace:

$\langle$"*Register*","*Check diesel engine*","*Test drive*","*Measure exhaust diesel*,"*Call customer*"" $\rangle$

Is this case involved in any decision point? Explain.



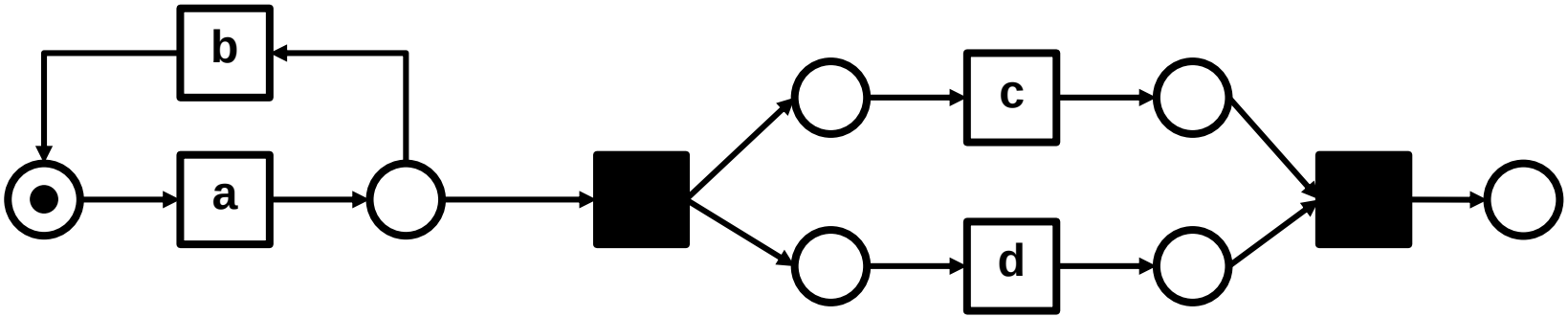
Exercise 4

Given the process model below, answer the following questions:

a) **What are the decisions?**

b) **Which type of table can you create? Event-based situation table, case-based situation table, or both?**

c) **In your table, what does each row represent?**



Decision Mining with Celonis

- Upload the files “event_table.csv” and “case_table.csv” into Celonis. Create a corresponding data model using the CASE ID to connect the activity table (“event_table.csv”) and the case table (“case_table.csv”).
- Create a new analysis using the newly created data model.

See our videos, former instructions, and the provided manual.

PQL documentation: https://docs.celonis.com/en/pql-function-library.html

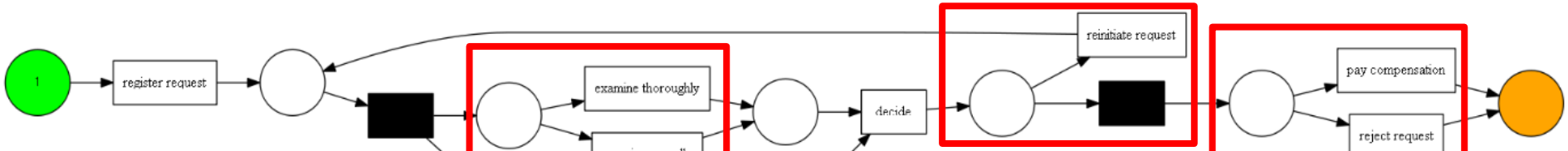
Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Case-based situation tables with PQL

- **Add a new OLAP component into your analysis. This table must contain the following variables:**
  - → **CASEID**
  - → **Resource of 1st activity (of case)**
  - → **Sum of costs of all events (of case)**
  - → **No. of “examine” activities (within case)**
  - → **No. of involved resources (in case)**
  - → **‘Yes’ if one resource executes more than one activity, ‘No’ otherwise**
  - → **No. of active cases (running cases) at start (of case)**

Decision points in the data

Decision point #1: possibly multiple decisions for the same case.
Decision point #2: only one decision for one case.
Decision point #3: possibly multiple decisions for the same case.



decision #1:
examine thoroughly
OR
examine casually

decision #3:
Reinitiate
OR
Not reinitiate

decision #2:
pay compensation
OR
reject request

In the following, we focus on decision points #1 and #2.

Case-based table for decision with PQL

- **Add a new OLAP component into your analysis. This table must contain the following variables:**
  - → **CASEID**
  - → **Resource of last activity (of case)**
  - → **Sum of costs of all events (of case)**
  - → **Decision: ‘pay’ if case ends with “pay compensation” and ‘reject’ otherwise.**

Case-based table for decision with PQL

- **Hide the CASEID column and export the table. Apply the Decision Tree classifier in RapidMiner using the Decision column as label.**

Event-based table for decision with PQL

- **Add a new OLAP component into your analysis. This table must contain the following variables:**
  - → **CASEID of event**
  - → **Resource of event**
  - → **No. running cases at event**
  - → **Sum of costs of all previous events (of same case)**
  - → **Decision: Activity of event**

- **Filter the component so that it only shows examine-events.**

- **Hide the CASEID column and export the table. Apply the Decision Tree classifier in RapidMiner using the Decision column as label.**