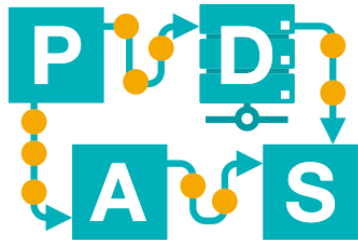


Decision Mining

Harry Beyel

BPI-I10



Chair of Process
and Data Science

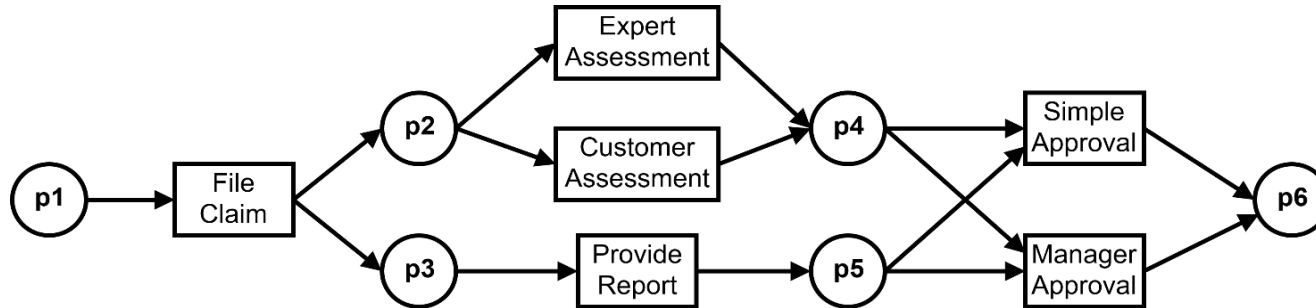
RWTHAACHEN
UNIVERSITY

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.

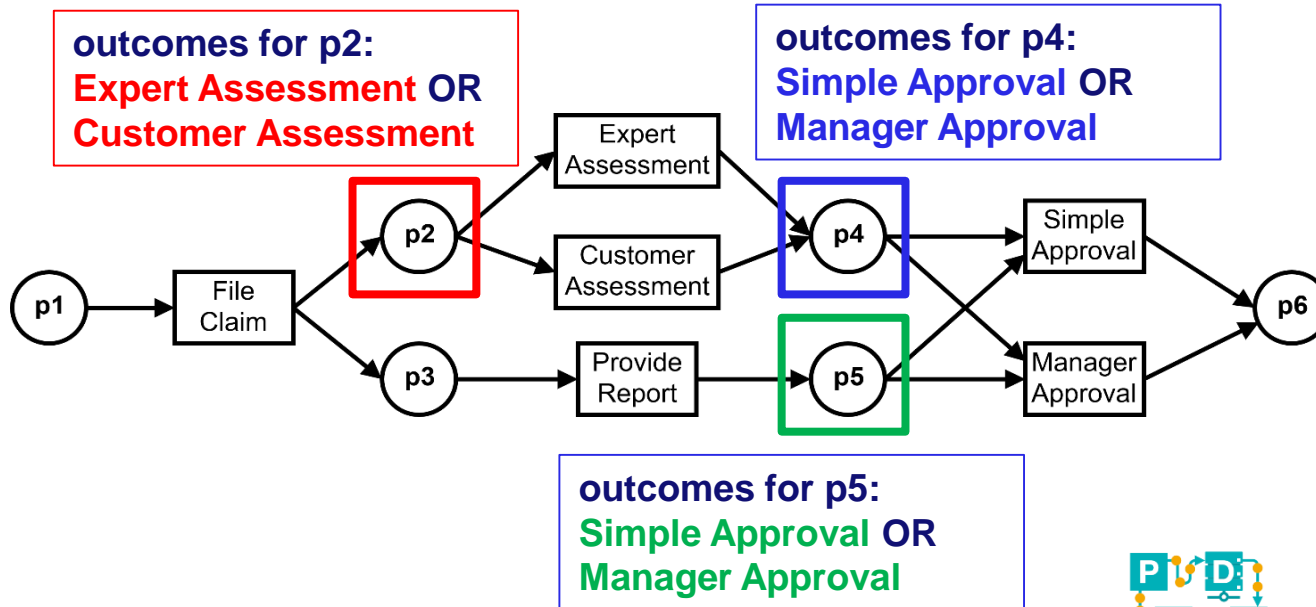
Exercise 1

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

a) Solution: Places p2, p4, and p5 are decision points.



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

b) Solution:

If you are given a non-trivial case table (e.g., in Celonis), then the columns of the case table are the case attributes.

If you are only given a flat event log (e.g., the so-called activity table in Celonis), case attributes are the event attributes that always have the same value for all events of the same case.

Here: *Coverage*, *Cost*, and *Type*

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

c) Outcomes:
Simple Approval OR
Manager Approval

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) Outcomes:
Simple Approval OR
Manager Approval

For each process run
(case) one single outcome:
We can use case-level
variables as predictor
variables.

Note: Every outcome is a
unique event. You can always
use event-level variables as
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

d) Solution:

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) Solution:

Case ID	Coverage	Cost	Type	Outcome
1	Extended	20000	Black	Manager
2	Standard	2400	Black	Simple
3	Extended	35000	Black	Manager
4	Standard	1900	Black	Simple
5	Extended	12000	Clean	Simple
6	Standard	15000	Black	Manager

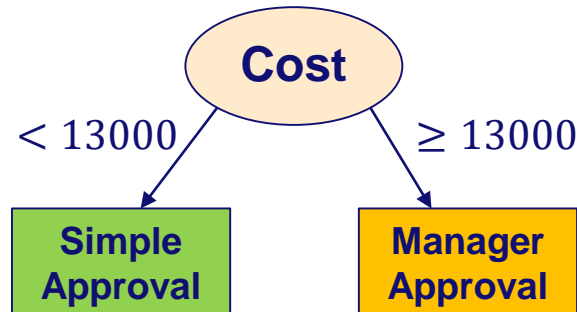
Exercise 1

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

Exercise 1

e) Solution:

Cost value: 1900 < 2400 < 12000 < 15000 < 20000 < 35000



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*.

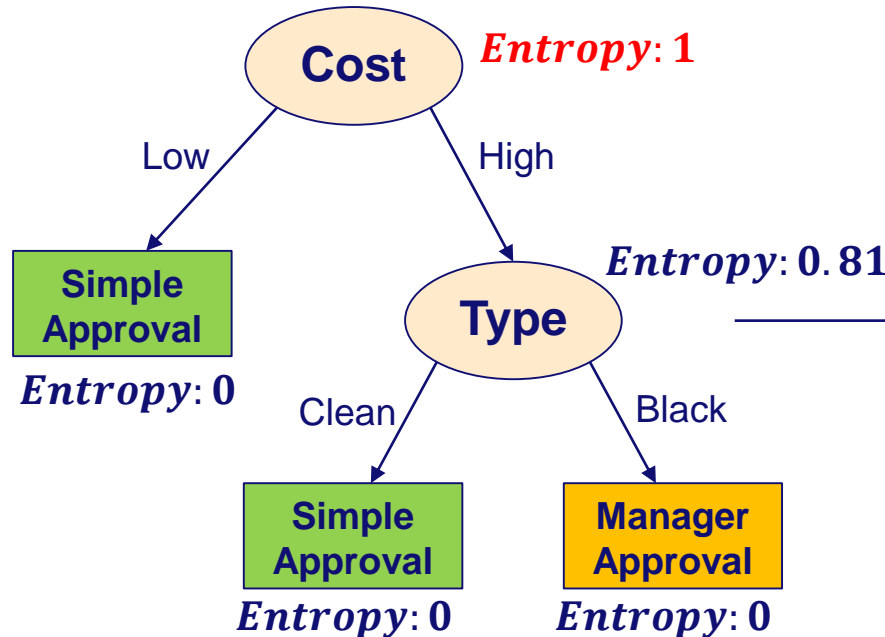
Exercise 1

f) Solution:

Case ID	Coverage	Cost	Type	Outcome
1	Extended	<i>High</i>	Black	Manager
2	Standard	<i>Low</i>	Black	Simple
3	Extended	<i>High</i>	Black	Manager
4	Standard	<i>Low</i>	Black	Simple
5	Extended	<i>High</i>	Clean	Simple
6	Standard	<i>High</i>	Black	Manager

Exercise 1

f) Solution: We select the predictor variables in the order that maximizes information gain in each step.



Weighted entropy:

$$\frac{2}{6} \times 0 + \frac{4}{6} \times 0.81 = \mathbf{0.54}$$

IG (for first split):

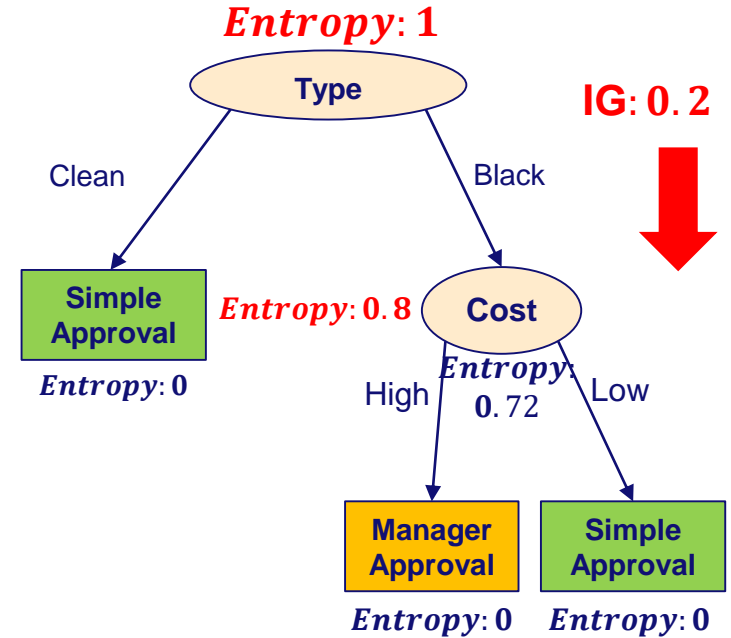
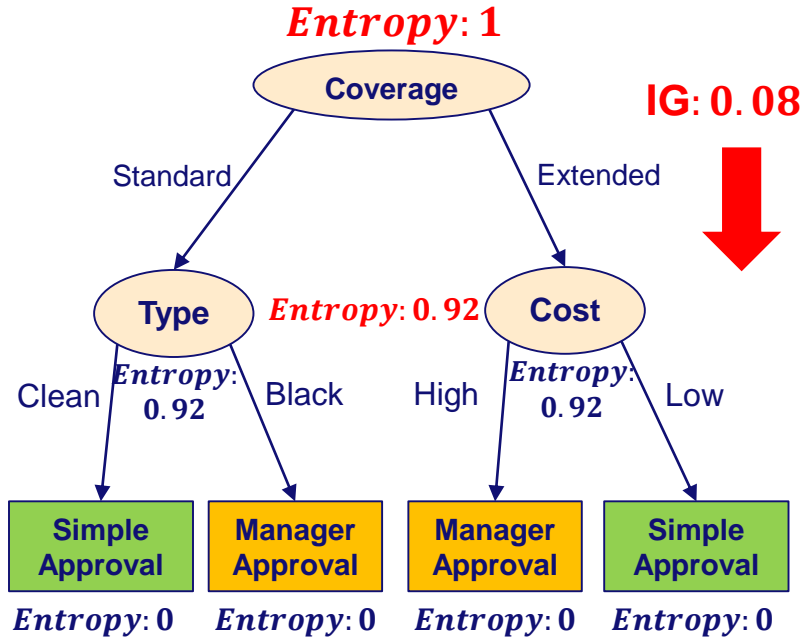
$$1 - 0.54 = \mathbf{0.46}$$

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*.

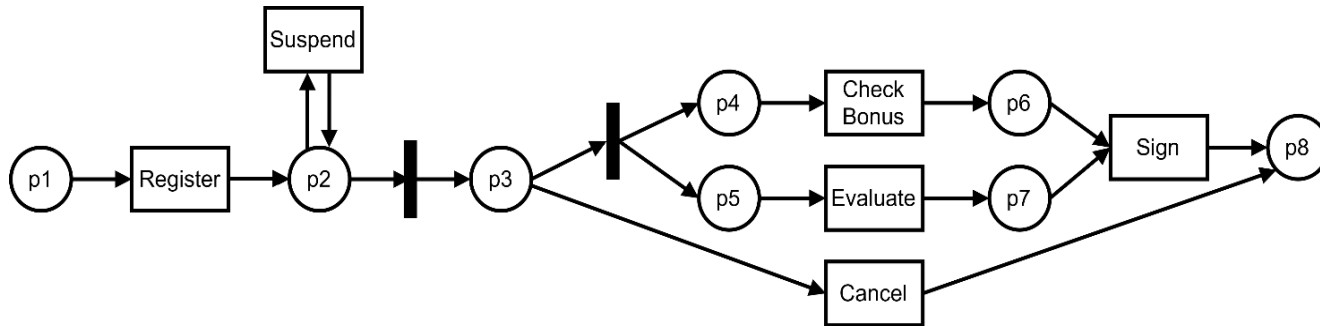
E.g., try constructing a decision tree for f) where the root attribute is *Coverage* and another one where the root attribute is *Type*.

Exercise 1



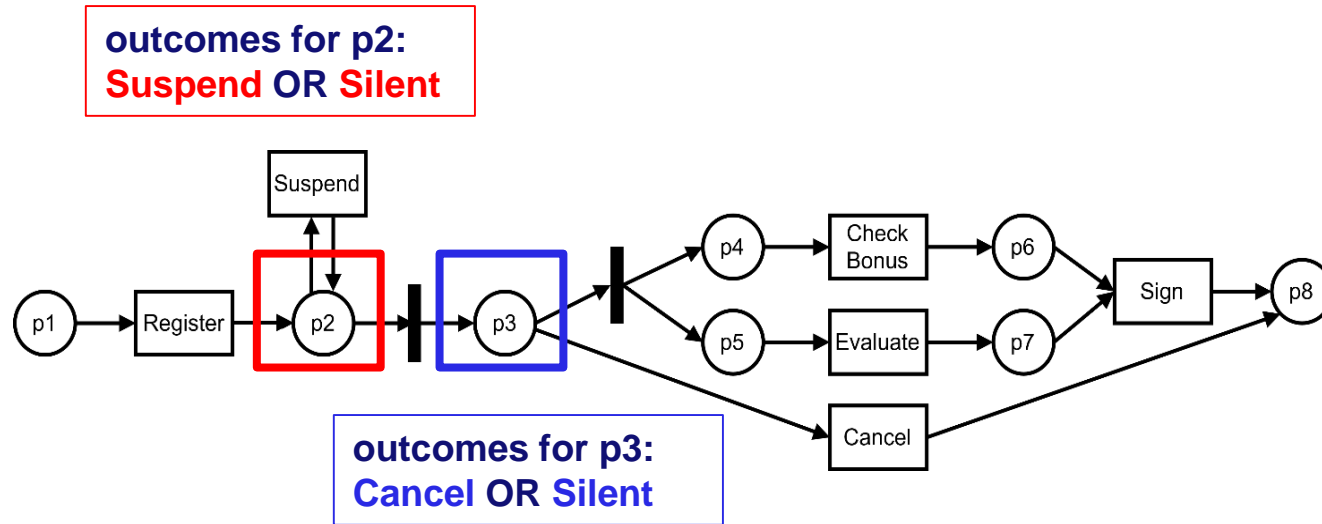
Exercise 2

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

a) Solution:



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

b) Solution:

Case ID	# suspensions	Resource of Register	Outcome
1	0	Mike	No Cancel
2	2	James	Cancel
3	1	James	No Cancel
4	2	Mike	No Cancel
5	3	James	Cancel
6	0	James	No Cancel

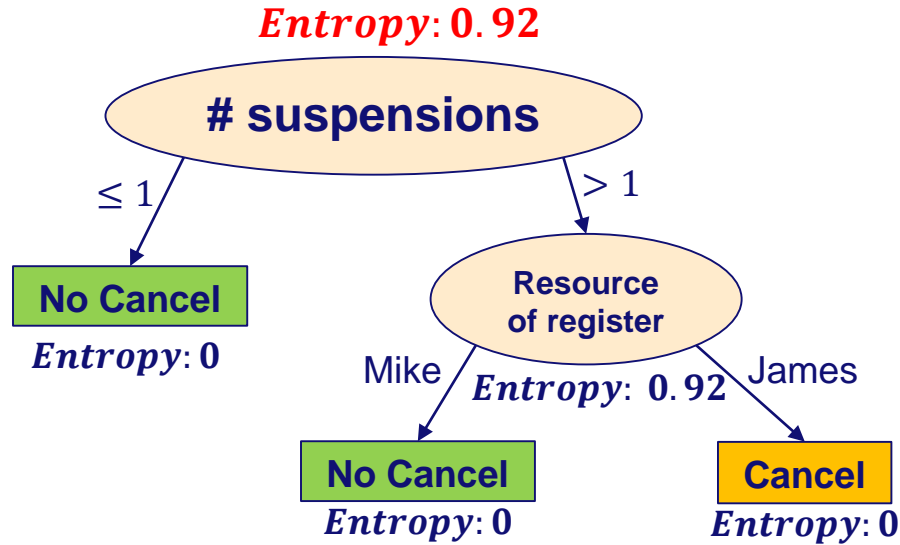
Case ID	Activity	Transition	Timestamp	Resource
1	Register	complete	08.01.2021	Mike
1	Check Bonus	complete	09.01.2021	John
1	Evaluate	complete	11.01.2021	Maria
1	Sign	complete	14.01.2021	Maria
2	Register	complete	09.01.2021	James
2	Suspend	complete	10.01.2021	Klaus
2	Suspend	complete	12.01.2021	Klaus
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

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.

Exercise 2

c) Solution:



Weighted entropy:

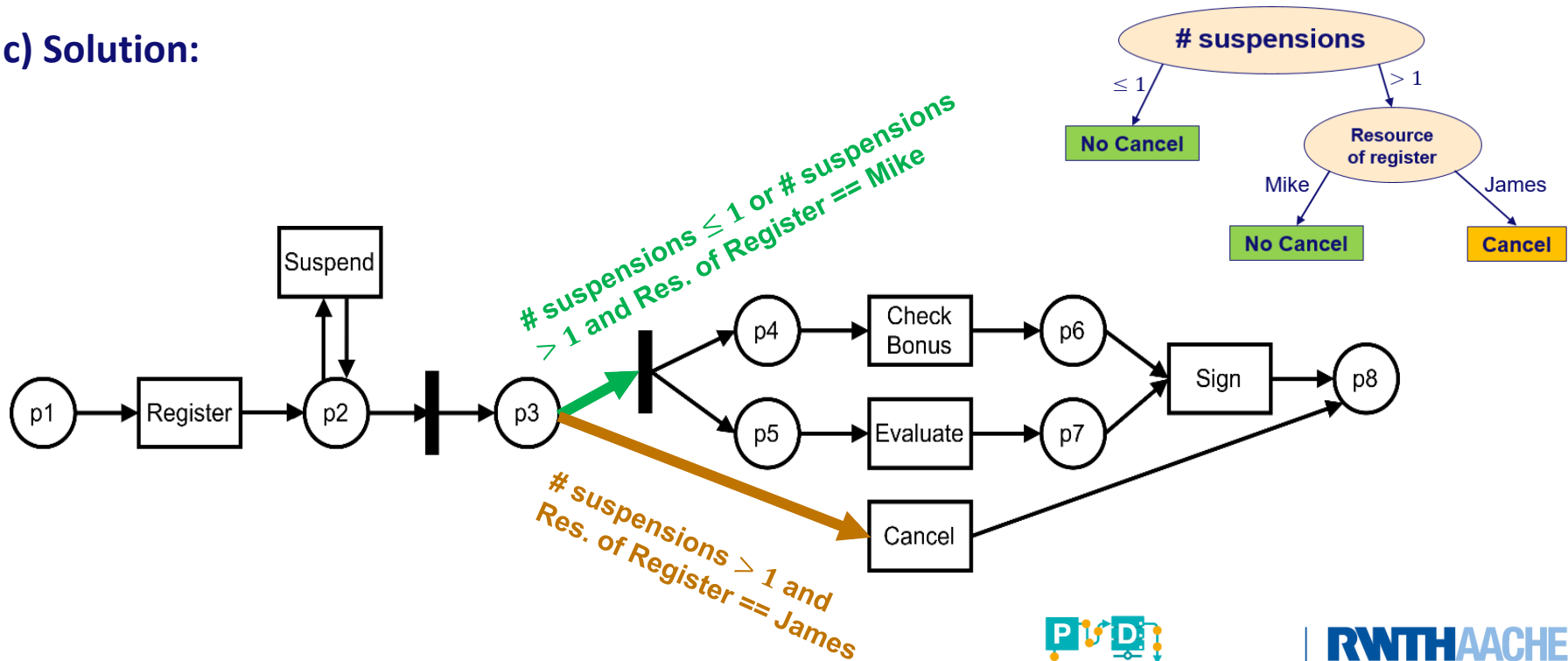
$$\frac{3}{6} \times 0 + \frac{3}{6} \times 0.92 = 0.46$$

IG (for first split):

$$0.92 - 0.46 = 0.46$$

Exercise 2

c) Solution:



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.

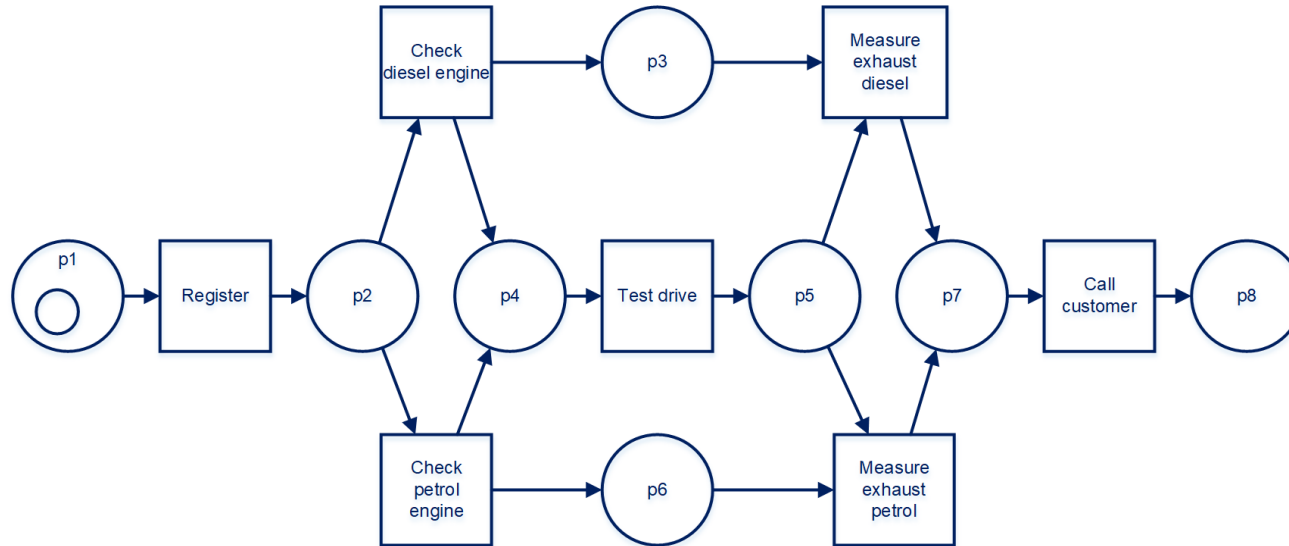
Decision Mining with Silent Transitions

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.

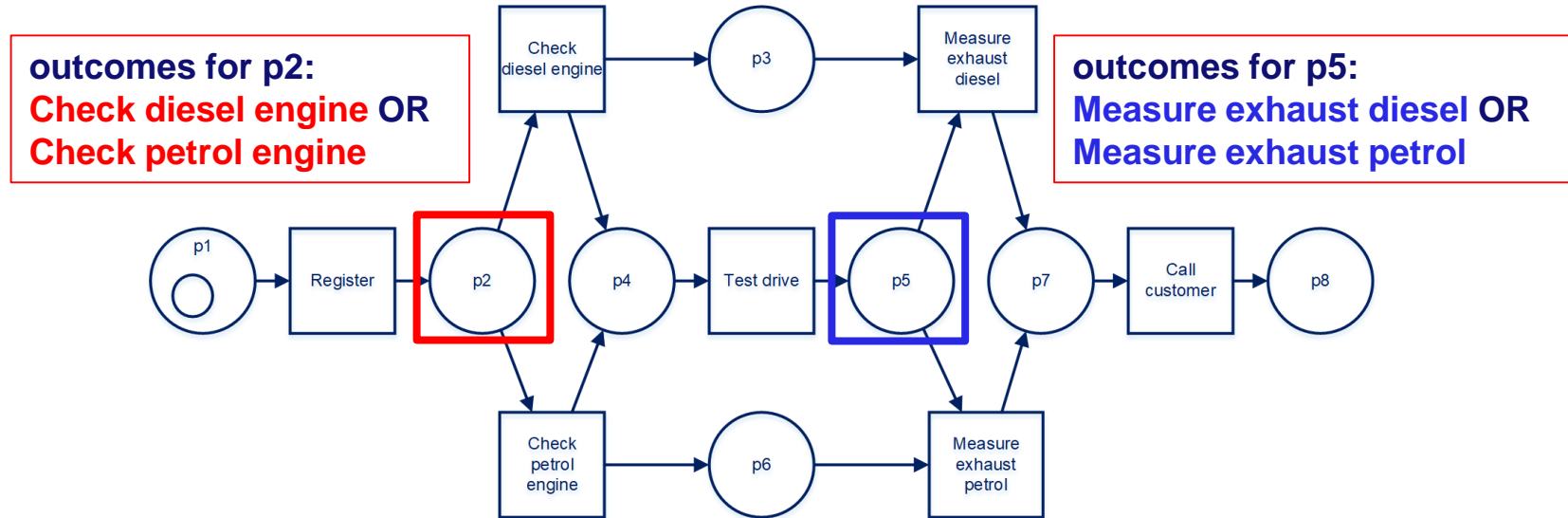
Exercise 3

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



Exercise 3

a) Solution:

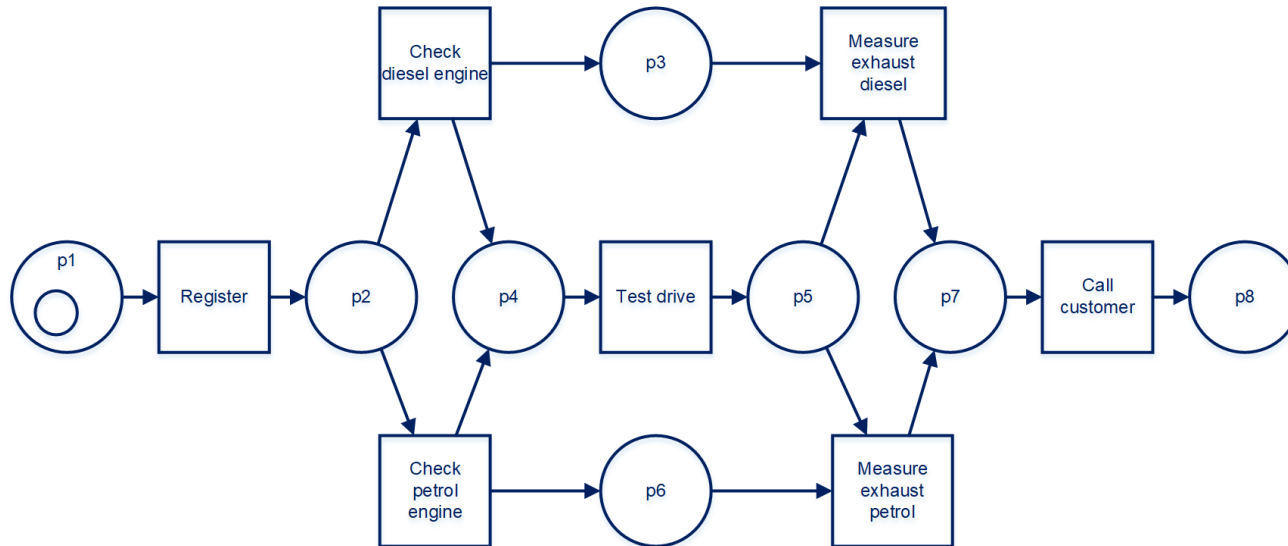


Exercise 3

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

⟨“Register”, “Check diesel engine”, “Test drive”, “Measure exhaust diesel”, “Call customer”⟩

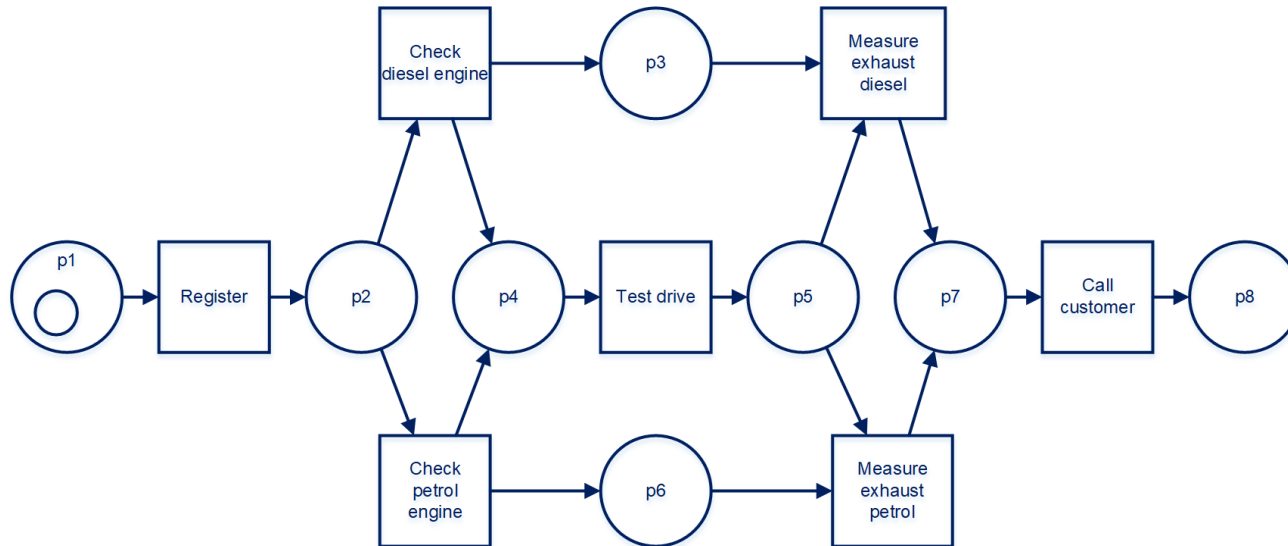
Is this case involved in any decision point? Explain.



Exercise 3

b) Solution:

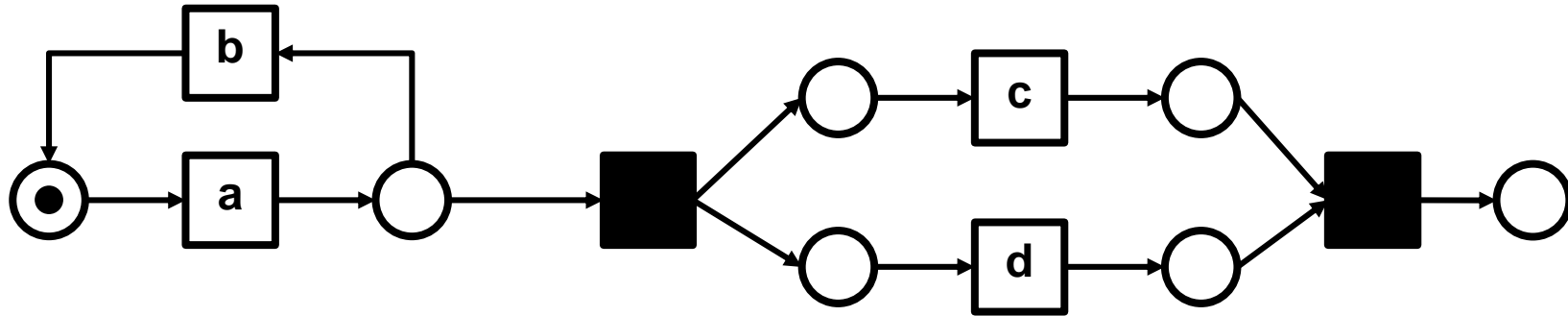
Trace *<“Register”, “Check diesel engine”, “Test drive”, “Measure exhaust diesel”, “Call customer”>* is involved in both decisions p2 and p5.



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?



Exercise 4

Solution:

a) What are the decisions?

b or *silent*. We recognize *silent* when either *c* or *d* gets executed after *a*.

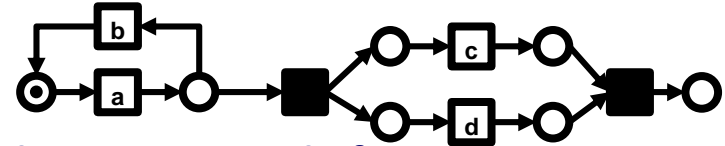
b) Which type of table can you create? Event-based situation table, case-based situation table, or both?

Decision in loop → event-based situation table

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

Each row represents a decision (*b* or *not-b*). *not-b* is denoted as *c* or *d*. Since every decision is related to the execution of *a*, we can denote a table as:

Case ID	Activity	Resource	...	Decision
1	a	X	...	b
1	a	Y	...	c or d
2	a	Y	...	c or d



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>

Case-based situation tables with PQL

- Add a new OLAP component into your analysis. This table must contain the following variables: **CASEID**, **CUSTOMER** and **AMOUNT**.
- 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)

Case-based situation tables with PQL

- Add a new OLAP component into your analysis. This table must contain the following variables: **CASEID**, **CUSTOMER** and **AMOUNT**.

CASEID:

"cases"."CASEID"

CUSTOMER:

"cases"."CUSTOMER"

AMOUNT:

"cases"."AMOUNT"

Whenever you want to add a new column to your table: Go to **DIMENSIONS** and select **Add**.

Go to **Custom dimension +** to open the PQL editor.

DIMENSIONS

Custom dimension +

Case-based situation tables with PQL

- Resulting table:

CASEID	CUSTOMER	AMOUNT
1	Silver	4000
2	Gold	500
3	Silver	2000
4	Silver	4000
5	Gold	8000
6	Gold	500
7	Silver	2000
8	Gold	500
9	Silver	4000
10	Gold	500
11	Gold	8000
12	Silver	4000
13	Silver	2000
14	Gold	500
15	Silver	4000
16	Gold	500

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)

Case-based situation tables with PQL

CASEID: `"cases"."CASEID"`

Resource of 1st activity: `PU_FIRST("cases", "events"."RESOURCE")`

Sum of costs of all events: `PU_SUM("cases", "events"."COST")`

No. of “examine” activities: `PU_COUNT("cases",
CASE WHEN "events"."ACTIVITY" IN
('examine casually', 'examine thoroughly')
THEN 1
ELSE NULL
END)`

Case-based situation tables with PQL

No. of involved resources :

```
PU_COUNT_DISTINCT("cases", "events"."RESOURCE")
```

‘Yes’ if one resource executes more than one activity, ‘No’ otherwise:

```
CASE WHEN PU_COUNT_DISTINCT("cases", "events"."RESOURCE")  
< PU_COUNT("cases", "events"."ACTIVITY")  
THEN 'Yes'  
ELSE 'No'  
END
```

Case-based situation tables with PQL

No. of active cases at (any) event's timestamp:

```
RUNNING_SUM(  
  CASE WHEN ACTIVITY_LAG("events"."ACTIVITY") IS NULL  
  THEN 1  
  WHEN ACTIVITY_LEAD("events"."ACTIVITY") IS NULL  
  THEN -1  
  ELSE 0  
  END,  
  ORDER BY ("events"."TIMESTAMP")  
)
```

Case-based situation tables with PQL

No. of active cases at (any) event's timestamp:

The screenshot illustrates the process of creating a formula in the PQL interface. On the left, a sidebar menu shows 'Saved formulas' highlighted. A red arrow points from this menu to a green 'Create formula' button. Another red arrow points from the button to the 'Saved formula' form on the right. The form includes fields for 'Name' (containing 'No. running cases at event'), 'Description', and 'Parameters' (with an 'Add parameter' button). Below these is a 'Template' section with a code editor containing the following PQL code:

```
RUNNING_SUM(  
  CASE WHEN ACTIVITY_LAG("events"."ACTIVITY") IS NULL  
  THEN 1  
  WHEN ACTIVITY_LEAD("events"."ACTIVITY") IS NULL  
  THEN -1  
  ELSE 0  
  END,  
  ORDER BY ("events"."TIMESTAMP")  
)
```

At the bottom of the form are 'Remove' and 'Duplicate' buttons.

We save this formula under a KPI("formula name") and can reuse it later without rewriting the code.

Case-based situation tables with PQL

No. of active cases (running cases) at start (of case):

```
PU_FIRST("cases", KPI("No. running cases at event"))
```

You can find and click on the pre-saved KPI on the left menu.

Case-based situation tables with PQL

- Resulting table:

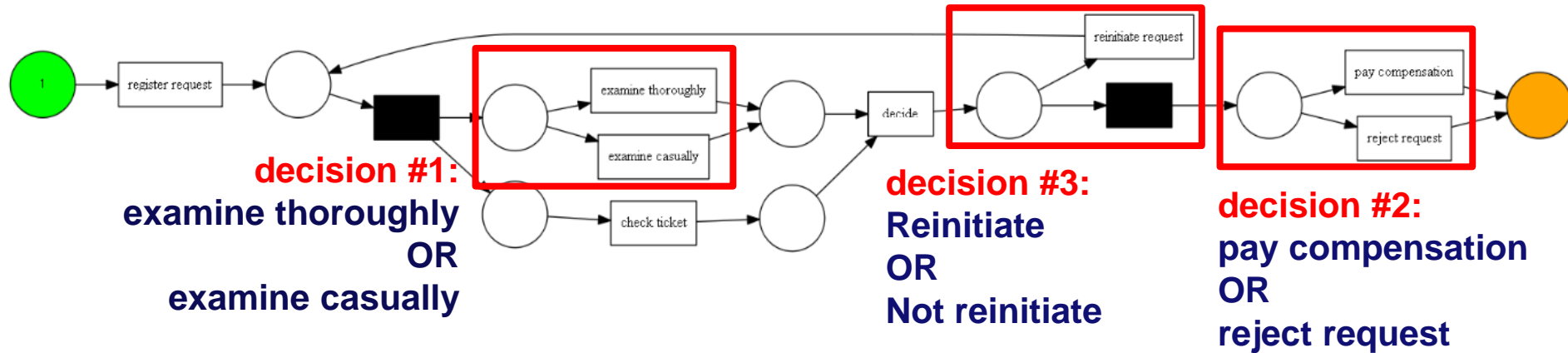
CASEID	Resource 1st event	Cost sum	# examinations	# involved resources	One resource multiple act...	No. running cases at start
1	Pete	950	1	4	Yes	1
2	Mike	950	1	4	Yes	51
3	Pete	1850	2	5	Yes	101
4	Pete	950	1	5	No	201
5	Ellen	2750	3	5	Yes	151
6	Mike	950	1	3	Yes	202
7	Pete	1850	2	5	Yes	102
8	Mike	950	1	4	Yes	52
9	Pete	950	1	4	Yes	2

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.



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

- Result:

CASEID	CUSTOMER	resource of last activity	Cost sum	decision
1	Silver	Pete	950	reject
2	Gold	Ellen	950	pay
3	Silver	Ellen	1850	pay
4	Silver	Ellen	950	reject
5	Gold	Mike	2750	reject
6	Gold	Mike	950	pay
7	Silver	Ellen	1850	pay
8	Gold	Ellen	950	pay
9	Silver	Pete	950	reject
10	Gold	Mike	950	pay
11	Gold	Mike	2750	reject
12	Silver	Ellen	950	reject
13	Silver	Ellen	1850	pay
14	Gold	Ellen	950	pay
15	Silver	Pete	950	reject

```
CASE WHEN  
  PU_LAST("cases",  
    "events"."ACTIVITY") =  
    'pay compensation'  
  THEN 'pay'  
  ELSE 'reject'  
END
```

```
PU_LAST("cases", "events"."RESOURCE")
```

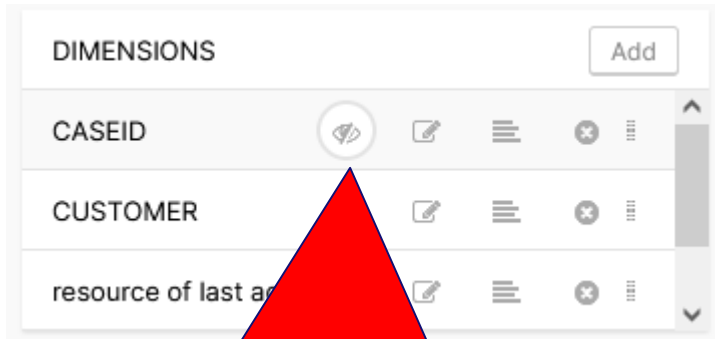


Chair of Process
and Data Science

RWTHAACHEN
UNIVERSITY

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.



Case-based situation table

CUSTOMER	resource of last activity	Cost sum	decision
Silver	Pete	950	reject
Gold	Ellen	950	pay
Silver	Ellen	1850	pay
Silver	Ellen	950	reject
Gold	Mike	2750	reject
Gold	Mike	950	pay
Silver	Ellen	1850	pay
Gold	Ellen	950	pay
Silver	Pete	950	reject

Click on the eye to remove the CASEID from the table (make it “invisible” without deleting it.)

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.

CUSTOMER	resource of last activity	Cost sum	decision
Silver	Pete	950	reject
Gold	Ellen	950	pay
Silver	Ellen	1850	pay
Silver	Ellen	950	reject
Gold	Mike	2750	reject
Gold	Mike	950	pay
Silver	Ellen	1850	pay
Gold	Ellen	950	pay
Silver	Pete	950	reject

Settings

Component filter

Export ...

Layers...

Copy component

Copy query

Export Data (XLSX)

Export Data (CSV)

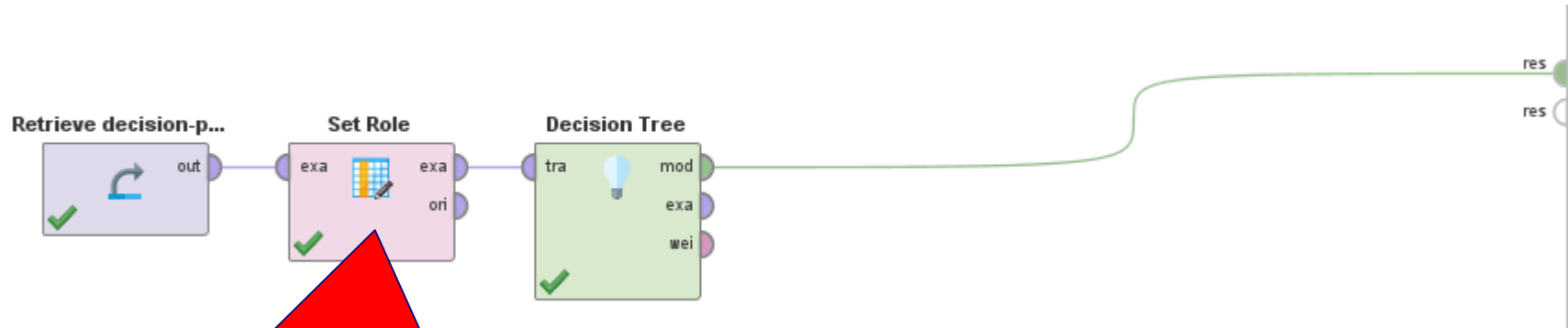
Export Cases (XLSX)

Export Cases (CSV)

**Download the table.
Import it into RapidMiner.**

Case-based table for decision with PQL

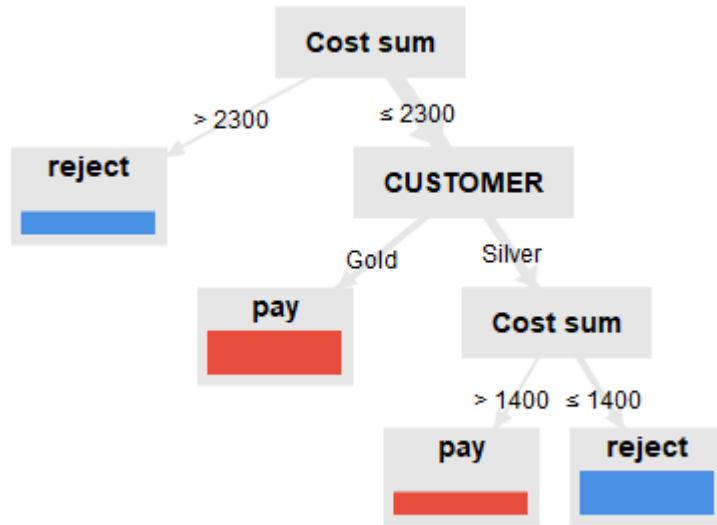
- Hide the CASEID column and export the table. Apply the Decision Tree classifier in RapidMiner using the Decision column as label.



Give the “Decision” column the role “label” in the configuration.

Case-based table for decision with PQL

- Result:



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.

Event-based table for decision with PQL

CASEID of event:

```
"events"."CASEID"
```

Resource of event

```
"events"."RESOURCE"
```

No. running cases at event

```
KPI("No. running cases at event")
```

This is the formula we saved!

Event-based table for decision with PQL

Sum of costs of all previous events (of same case)

```
RUNNING_SUM(  
  CASE WHEN ACTIVITY_LEAD ( "events"."ACTIVITY") IS NULL  
  THEN 0  
  ELSE "events"."COST"  
  END,  
  ORDER BY ("events"."TIMESTAMP"),  
  PARTITION BY ("events"."CASEID")  
)
```

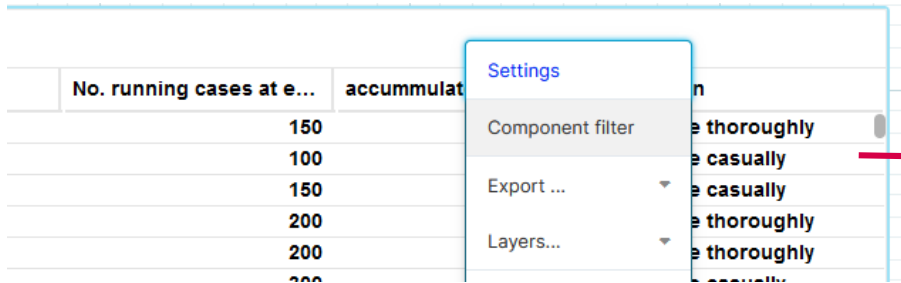

Event-based table for decision with PQL

→ Decision: Activity of event

"events"."ACTIVITY"

We are only interested in events that indicate choices: events with activity “examine casually” or “examine thoroughly”.

We filter the table so that it shows only the rows with those activities.



No. running cases at e...	accumulat
150	
100	
150	
200	
200	
200	

FILTER

```
"events"."ACTIVITY" IN  
( 'examine casually',  
  'examine thoroughly' )
```

Event-based table for decision with PQL

- Result:

Event-based situation table				
CASEID	RESOURCE	No. running cases at e...	accumulated costs	Decision
1	Sue	150	450	examine thoroughly
2	Sean	100	550	examine casually
3	Mike	150	450	examine casually
3	Sean	200	1350	examine thoroughly
4	Sean	200	550	examine thoroughly
5	Mike	300	450	examine casually
5	Mike	50	1450	examine casually
5	Sue	50	2250	examine casually
6	Ellen	300	450	examine casually
7	Mike	150	450	examine casually
7	Sean	200	1350	examine thoroughly
8	Sean	100	550	examine casually
9	Sue	150	450	examine thoroughly

Event-based table for decision with PQL

- Result (after hiding the CASEID column):

Event-based situation table			
RESOURCE	No. running cases at event	accumulated costs	Decision
Sue	150	450	examine thoroughly
Sean	100	550	examine casually
Mike	150	450	examine casually
Sean	200	1350	examine thoroughly
Sean	200	550	examine thoroughly
Mike	300	450	examine casually
Mike	50	1450	examine casually
Sue	50	2250	examine casually
Ellen	300	450	examine casually
Mike	150	450	examine casually
Sean	200	1350	examine thoroughly
Sean	100	550	examine casually
Sue	150	450	examine thoroughly

Event-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.



Give the “Decision” column the role “label” in the configuration.

Event-based table for decision with PQL

- Result:

