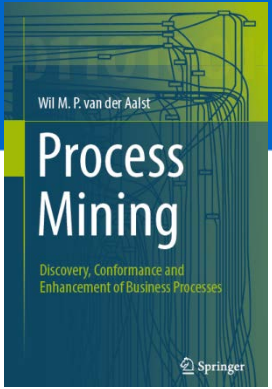
Process Mining: Data Science in Action

Mining Decision Points

prof.dr.ir. Wil van der Aalst

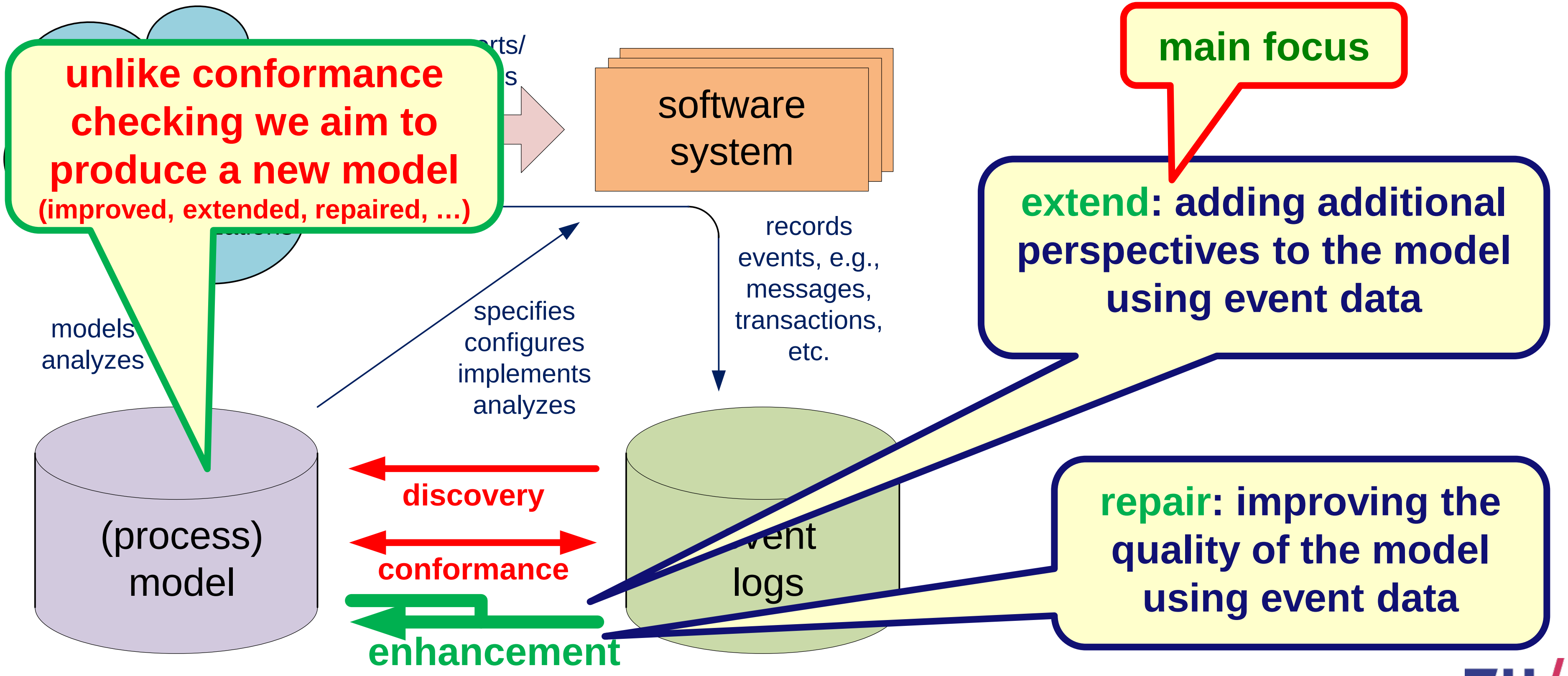
www.processmining.org

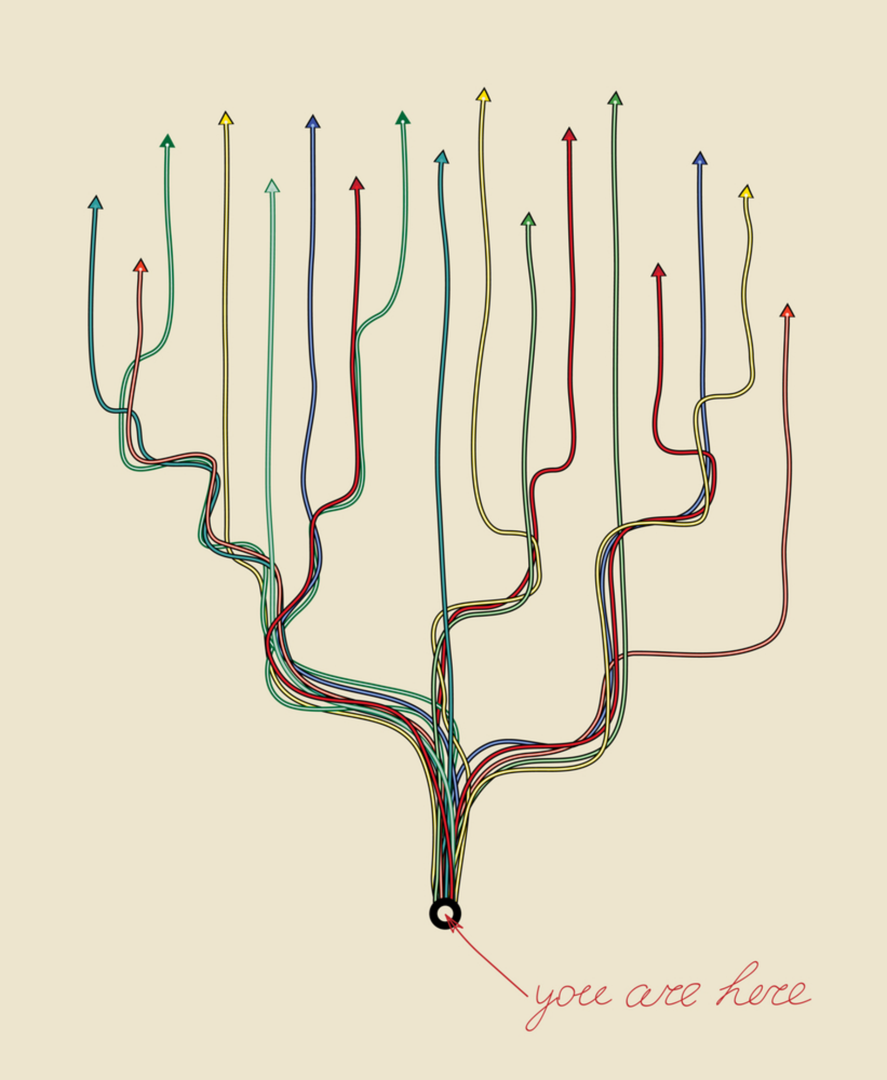


TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

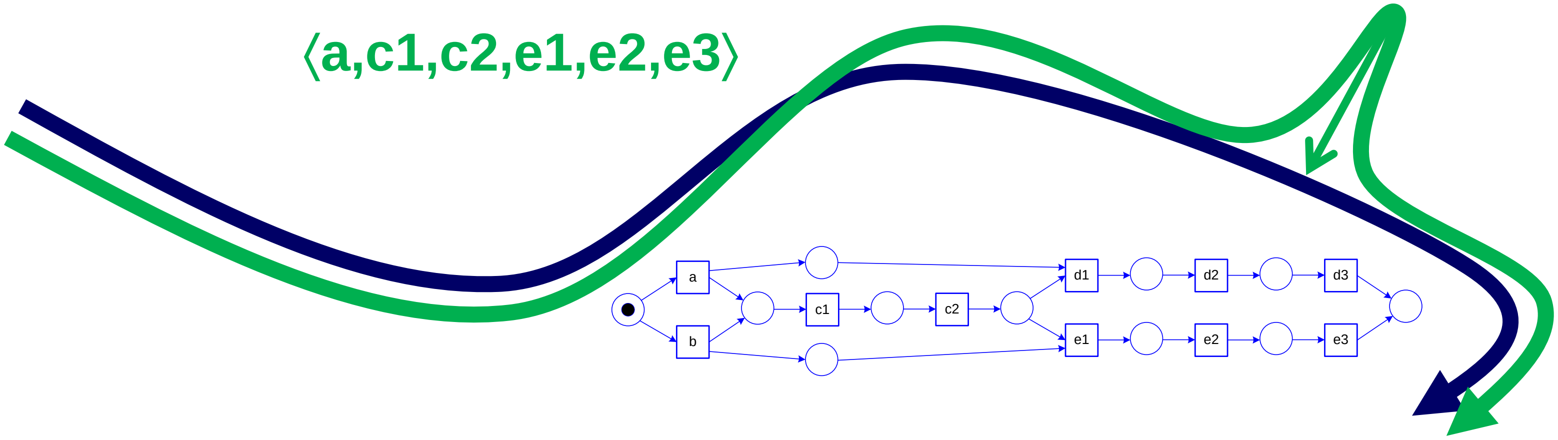
Enhancement: Extension and Repair



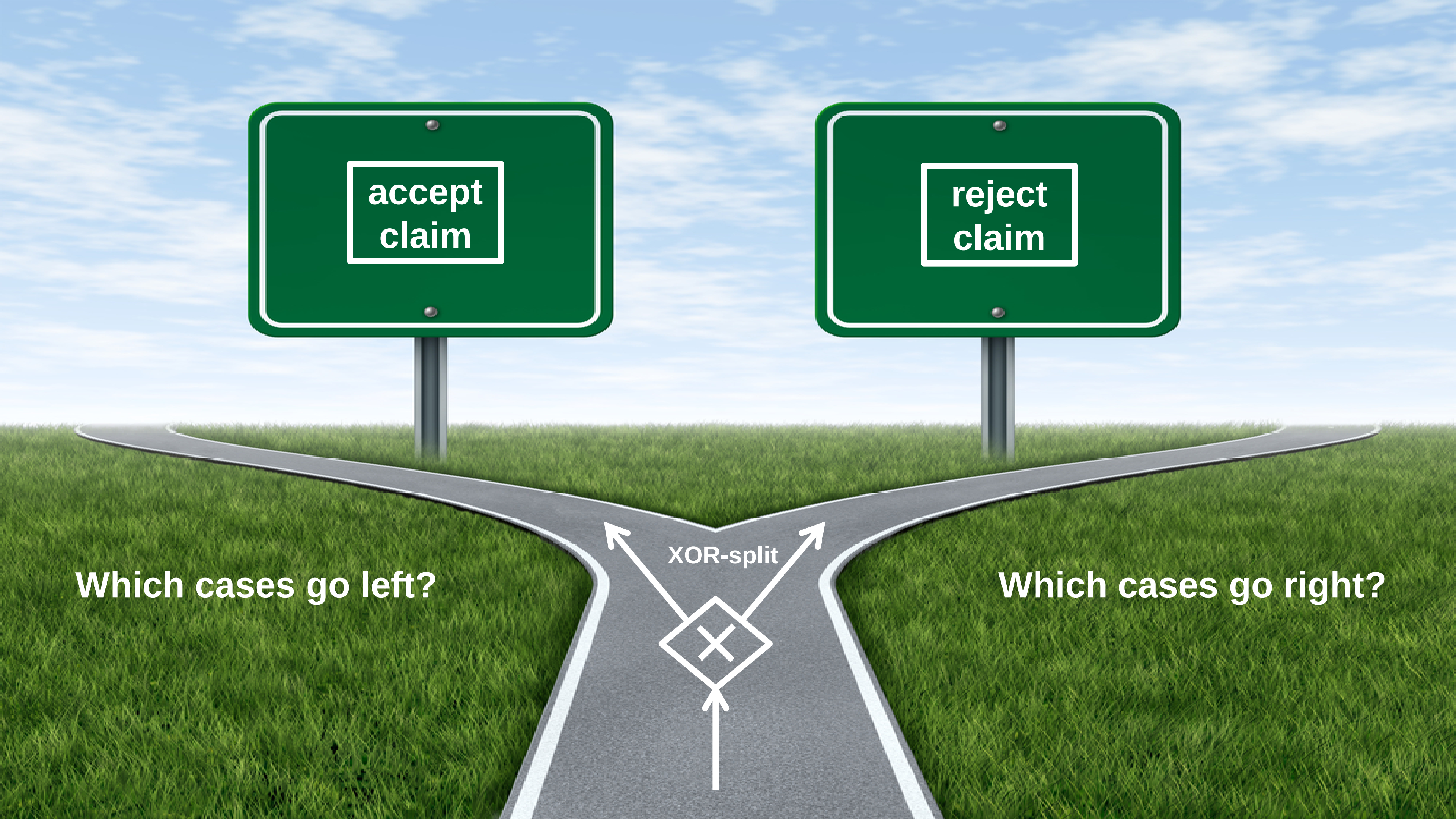


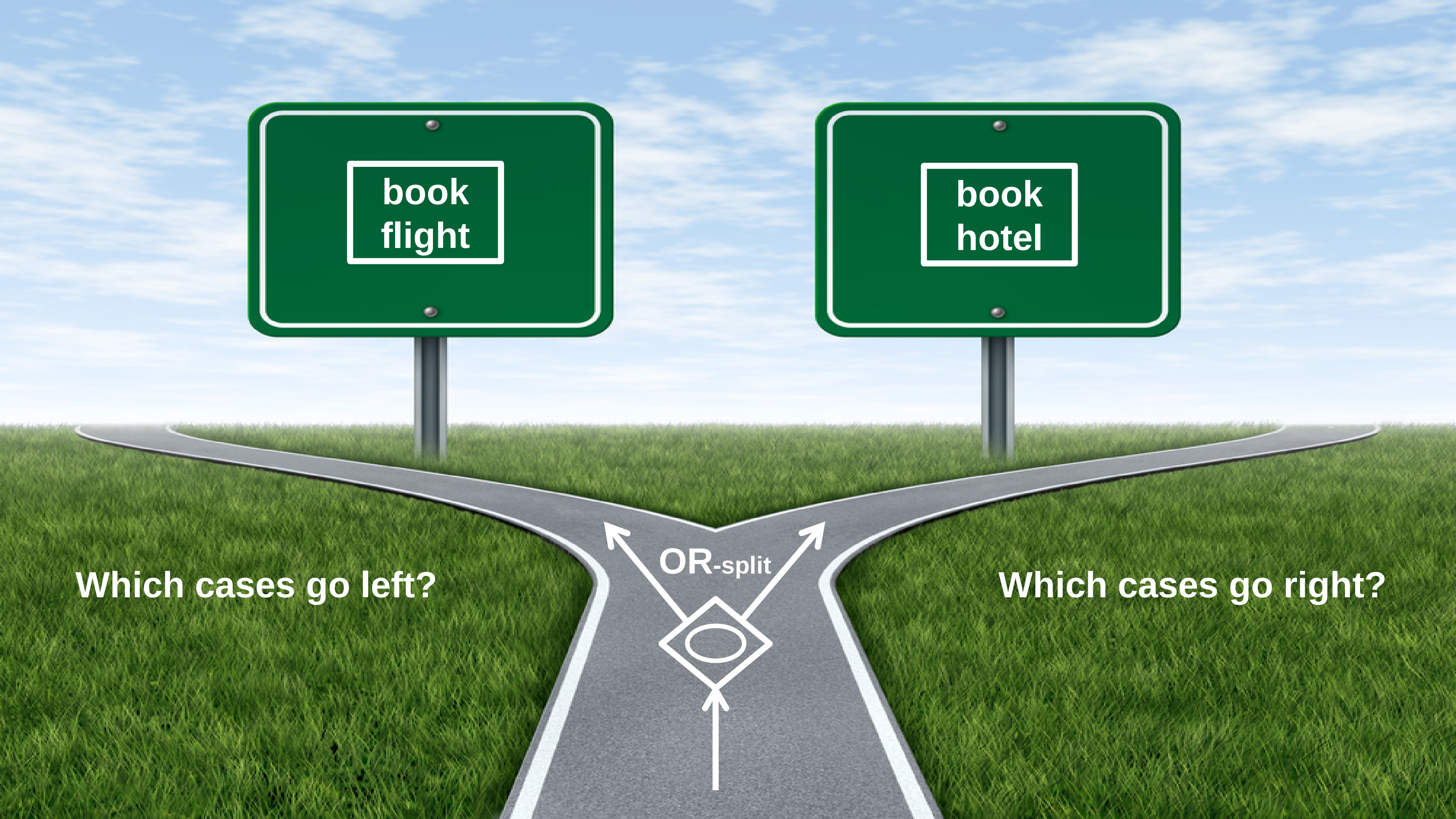
Mining Decision Points

- **Input:**
  - **event log**
  - **process model**
- **Assumption: Log and model have been aligned.**
  - **Mapping of activity names in log and model.**
  - **Every trace can be related to a path through the model.**



a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

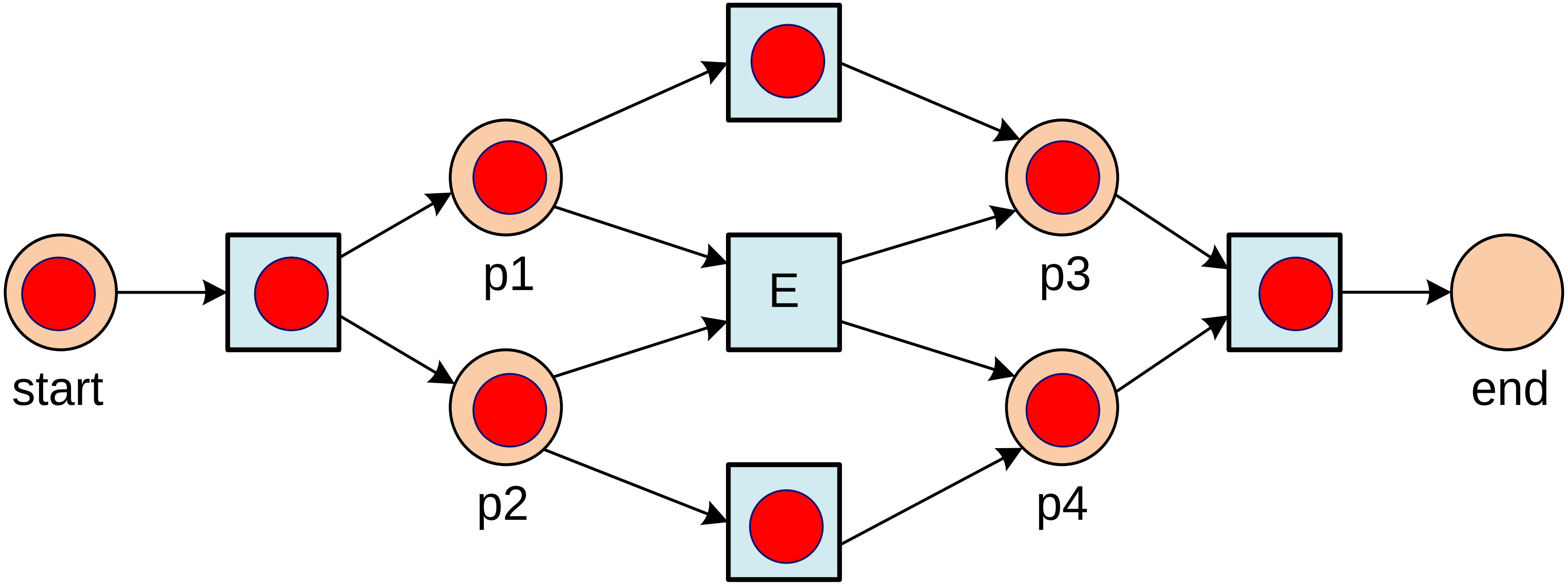

accept
claim
reject
claim
XOR-split
Which cases go left?
Which cases go right?



book
flight
book
hotel
OR-split
Which cases go left?
Which cases go right?

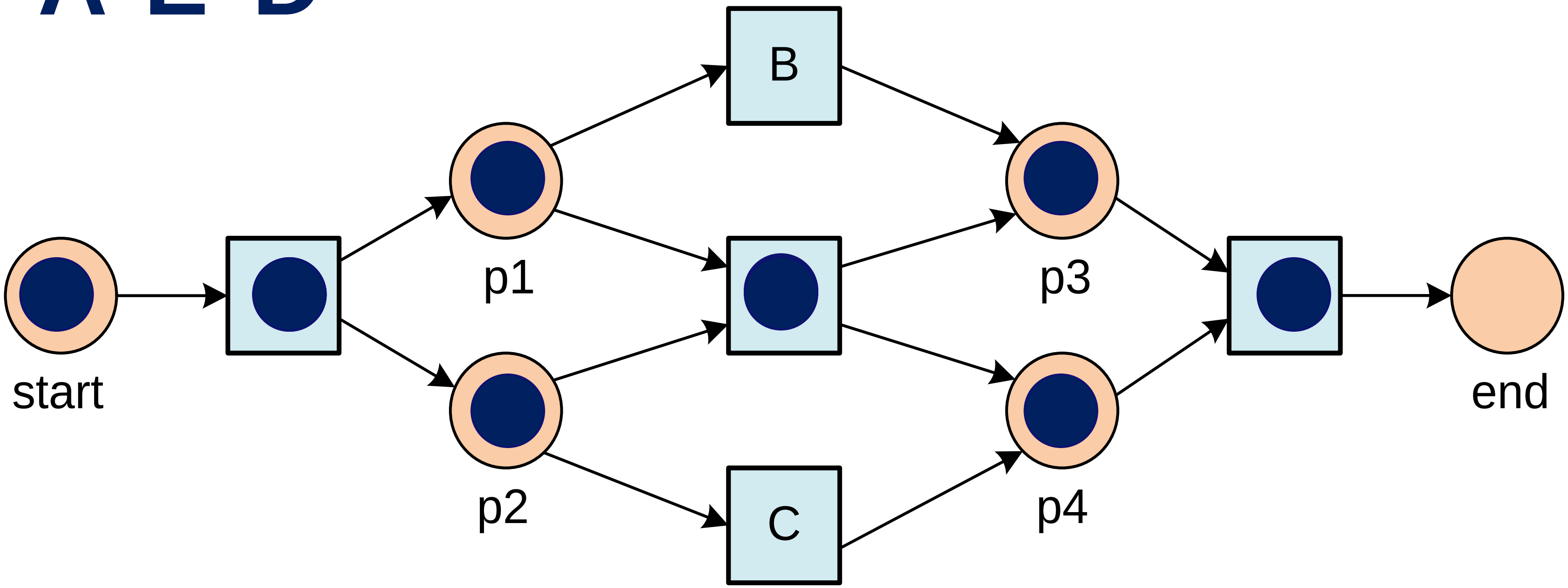
Decision mining: “Red” cases

A B C D



Decision mining: “Blue” cases

A E D

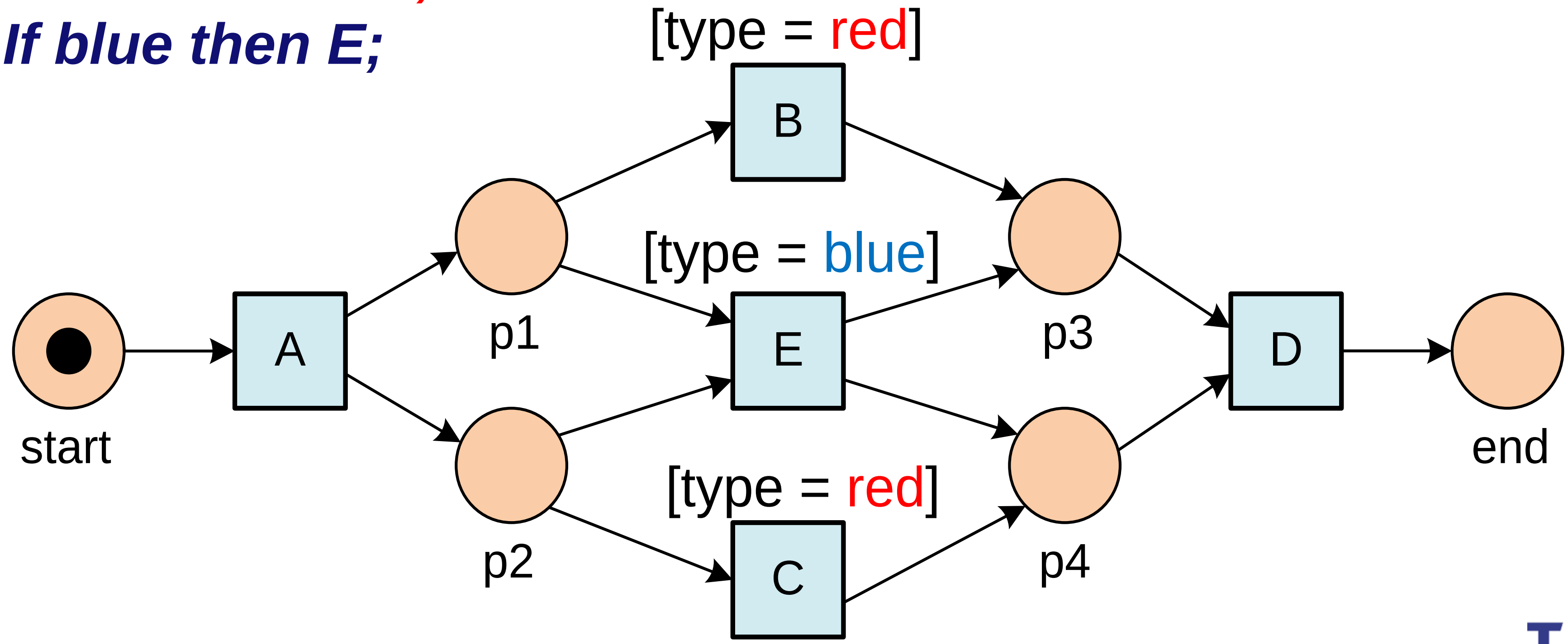


Guards ensure that the right path is taken

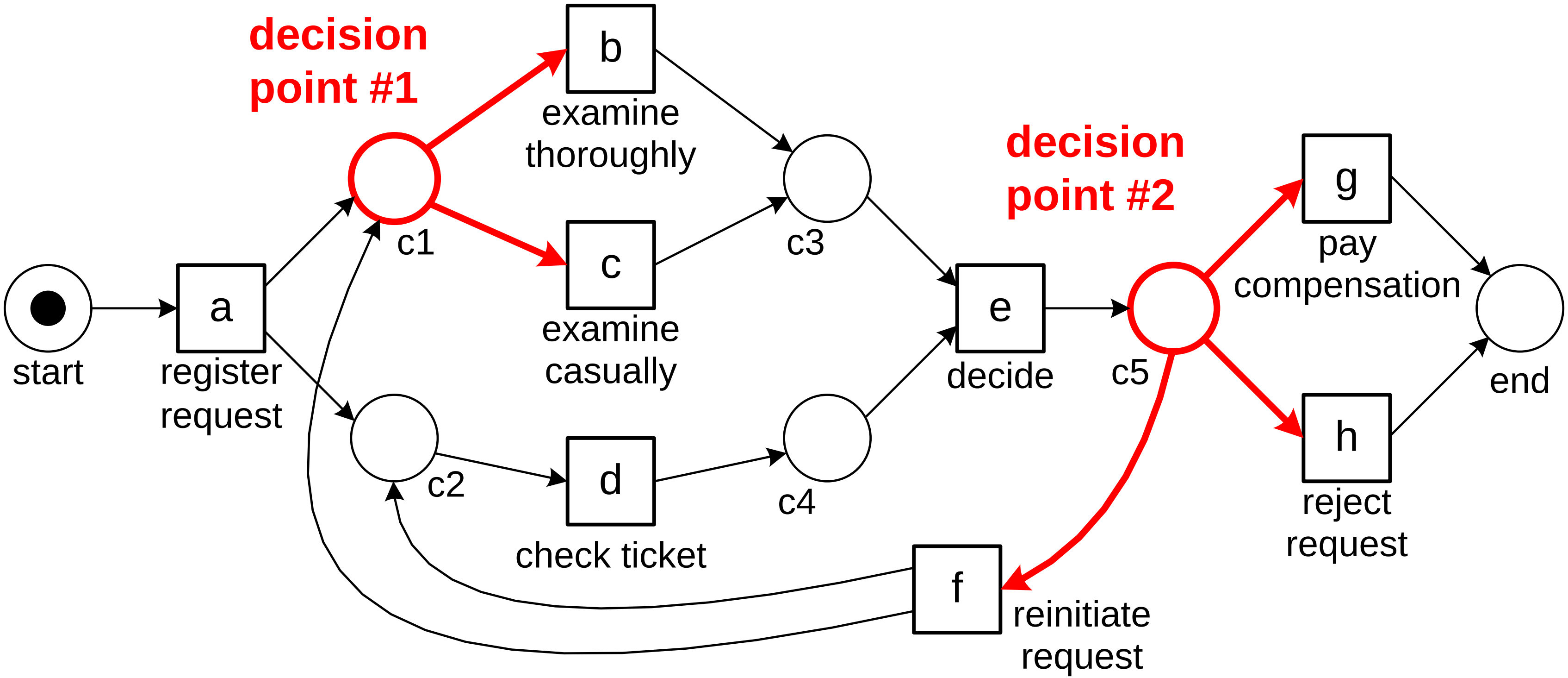
(assuming cases have a data attribute *type* having value blue or red)

If red then B+C;

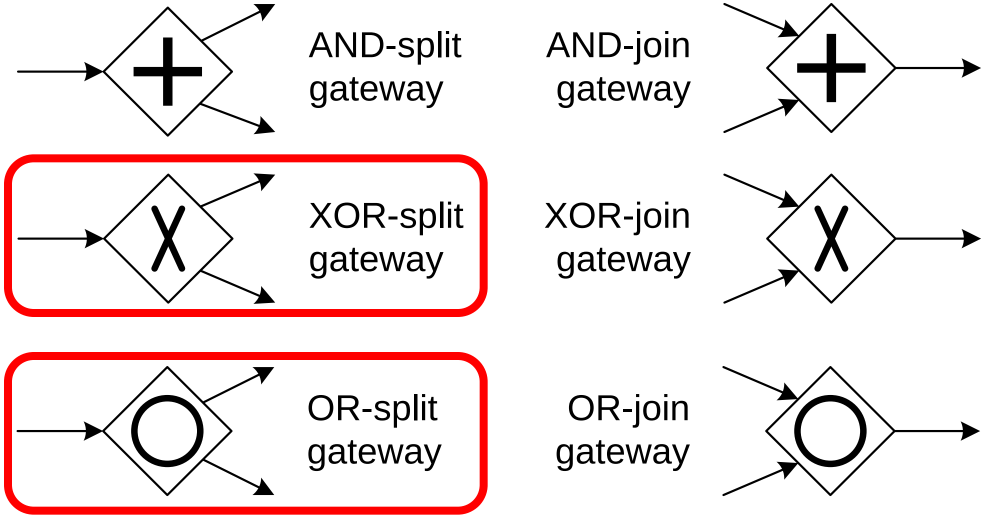
If blue then E;

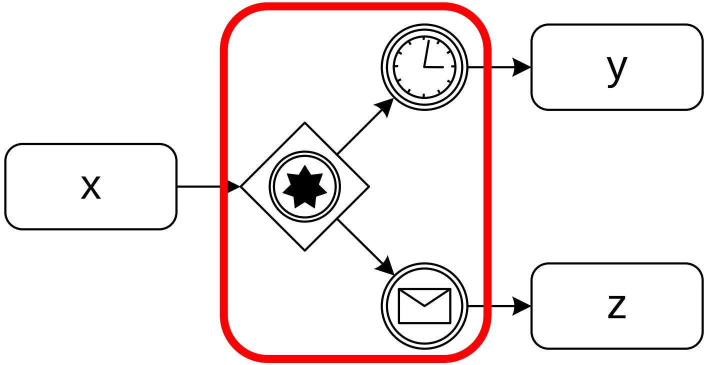


Places with multiple output arcs form decision points



Decision points in BPMN





event-based XOR-split gateway
(deferred choice pattern)


data
mining
process
mining

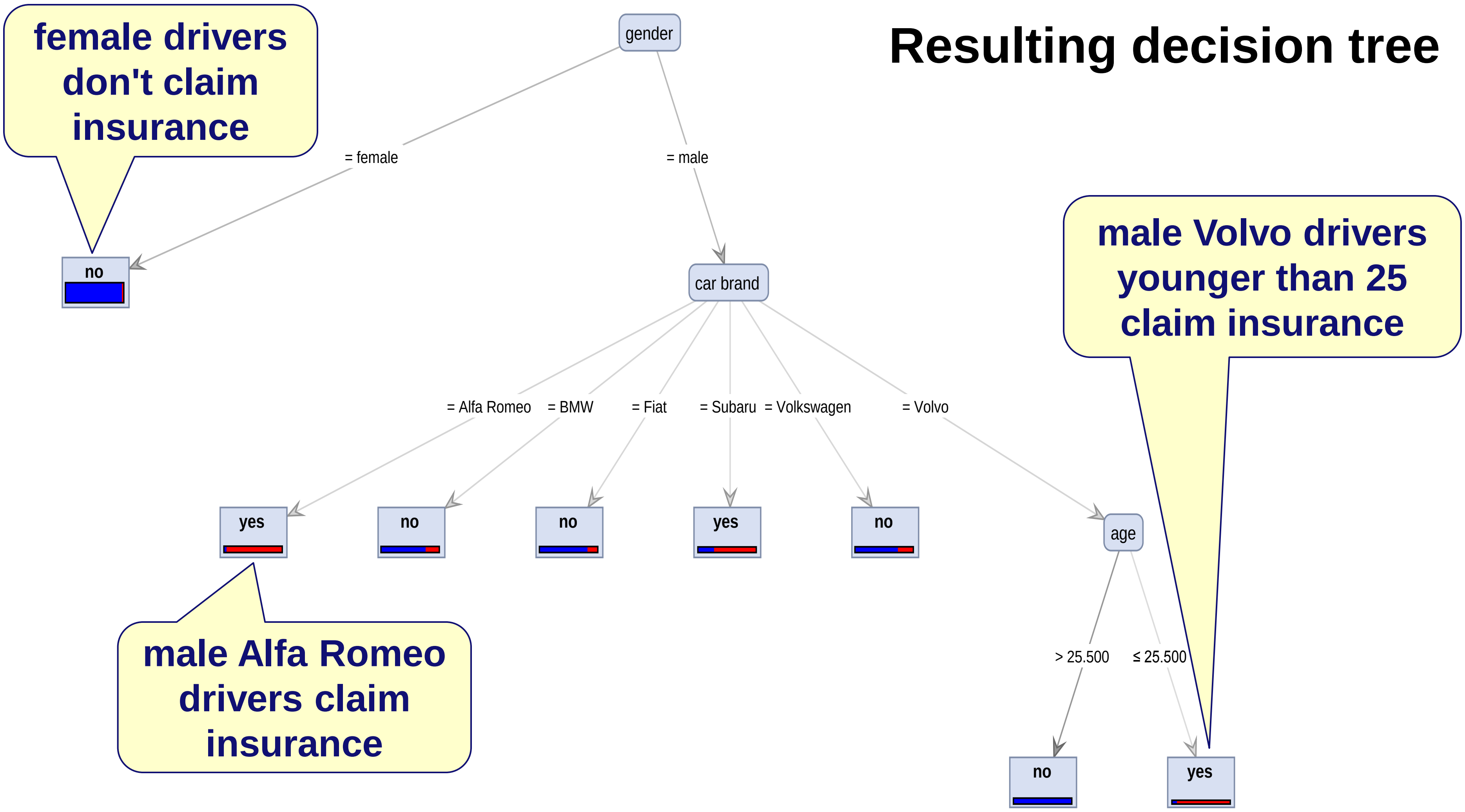
Remember: Classification using decision trees

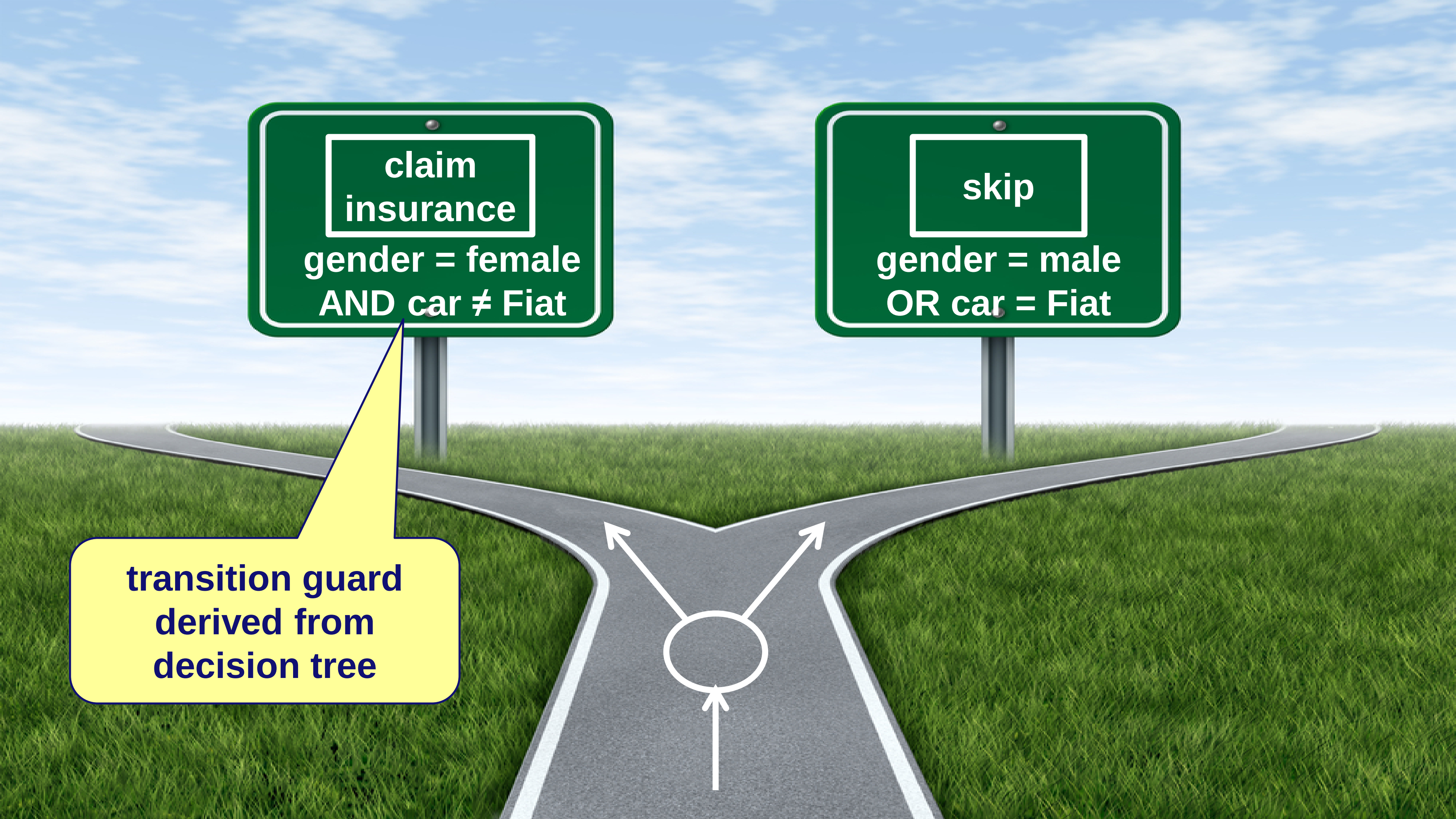
gender	age	smoker	car brand	claim
female	47	yes	Volvo	no
male	31	no	Alfa Romeo	yes
male	59	no	Alfa Romeo	yes
male	28	no	Fiat	no
male	44	no	BMW	no
female	27	no	Fiat	no
male	29	no	Subaru	no
male	44	yes	Subaru	yes
male	39	no	BMW	no
male	35	no	Subaru	yes

- **Response variable (dependent variable): claim (yes/no).**
- **Predictor variables (independent variables): gender, age, smoker, car brand.**

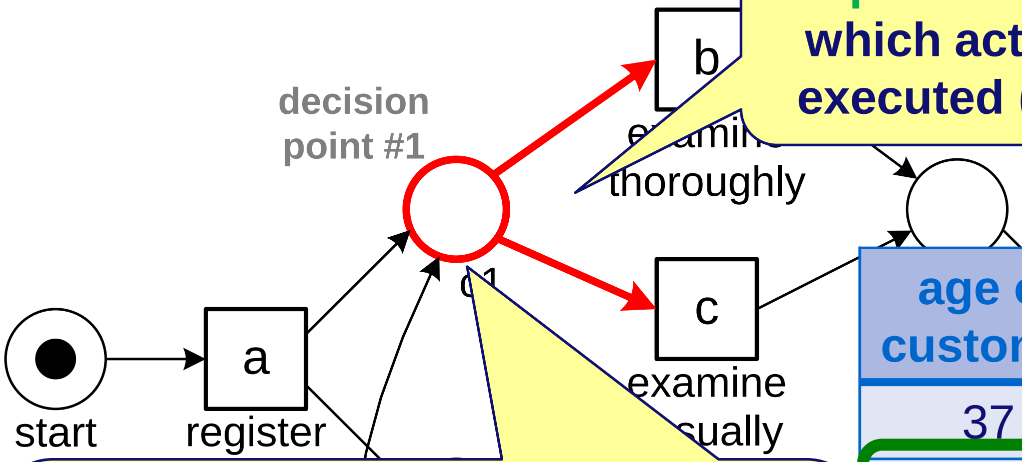
Goal: explain response variable in terms of relevant predictor variables.

Resulting decision tree




claim insurance
gender = female AND car ≠ Fiat
skip
gender = male OR car = Fiat
transition guard derived from decision tree

Creating a classification problem



response variable: which activity is executed (b or c)

instance = a decision and its context

age of customer	resource exec. a	ather	activity executed
37	Pete	sunny	b
63	John	sunny	c
39	Pete	rain	b
25	Mary	sunny	b
29	Sue	storm	c
…	…	…	…

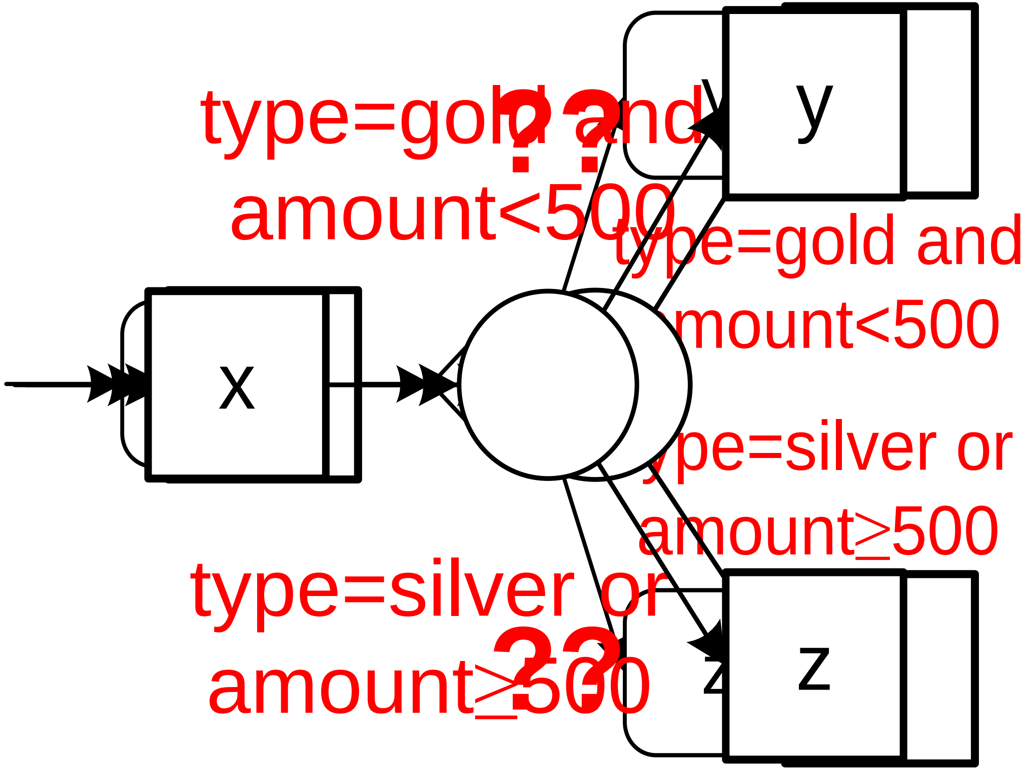
predictor variables: anything related to the context of this decision, e.g., the age of the client, the resource that executed the previous activity, the weather, etc.

Learning an XOR-split

type	region	amount	activity
silver	south	687.70	z
gold	south	987.30	z
silver	north	378.80	z
gold	south	314.50	y
silver	north	537.70	z
silver	west	158.70	z
gold	east	344.50	y
...	...	...	...

predictor variables:
- **type = silver**
- **region = south**
- **amount = 687.70**

response variable
- **activity = z**



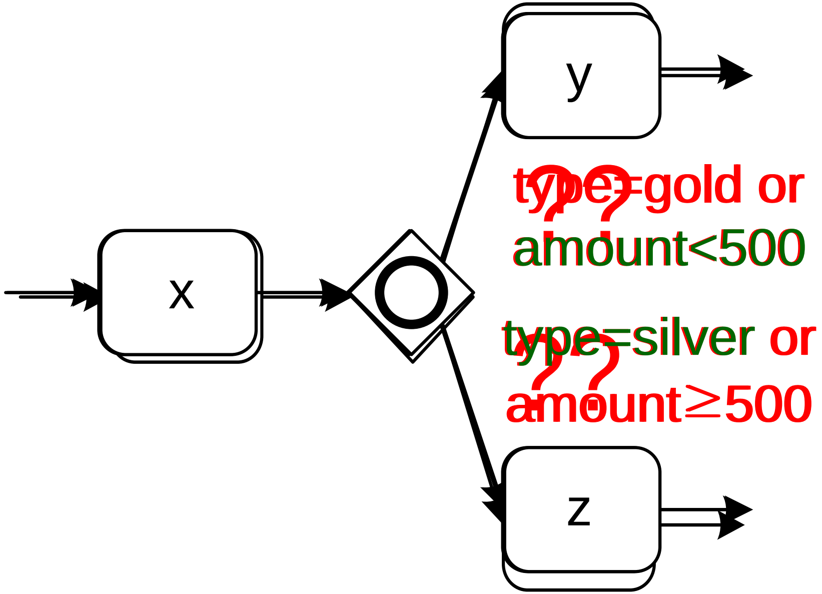
For any process notation, e.g., learning a BPMN XOR-split gateway.

Learning an OR-split

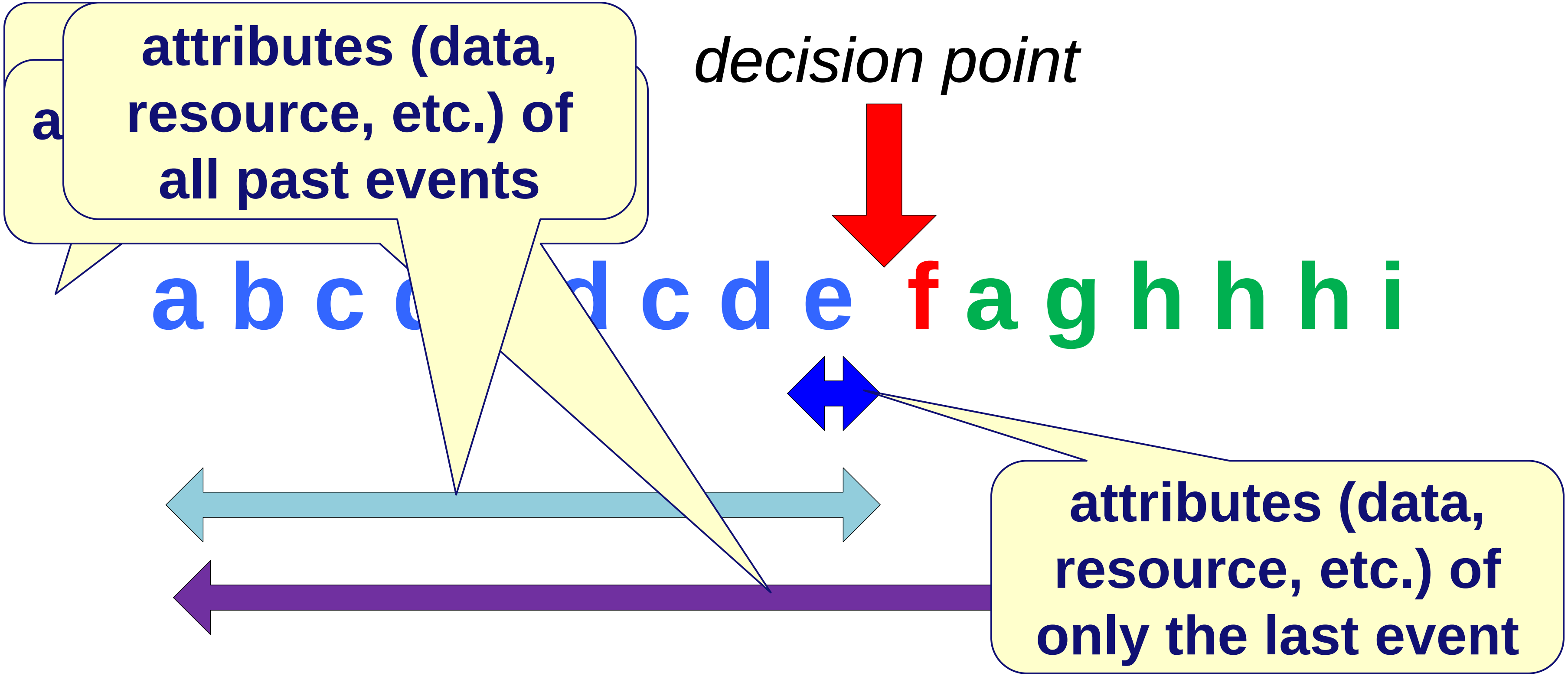
type	region	amount	activity
gold	south	987.30	y and z
silve			d z
gold			y
silve			z
silve			d z
silve			d z
gold	north	488.50	just y
silver	west	443.20	y and z
silver	south	673.70	just z
…	…	…	…

predictor variables:
- **type = silver**
- **region = west**
- **amount = 443.20**

response variable
- **activity = y and z**

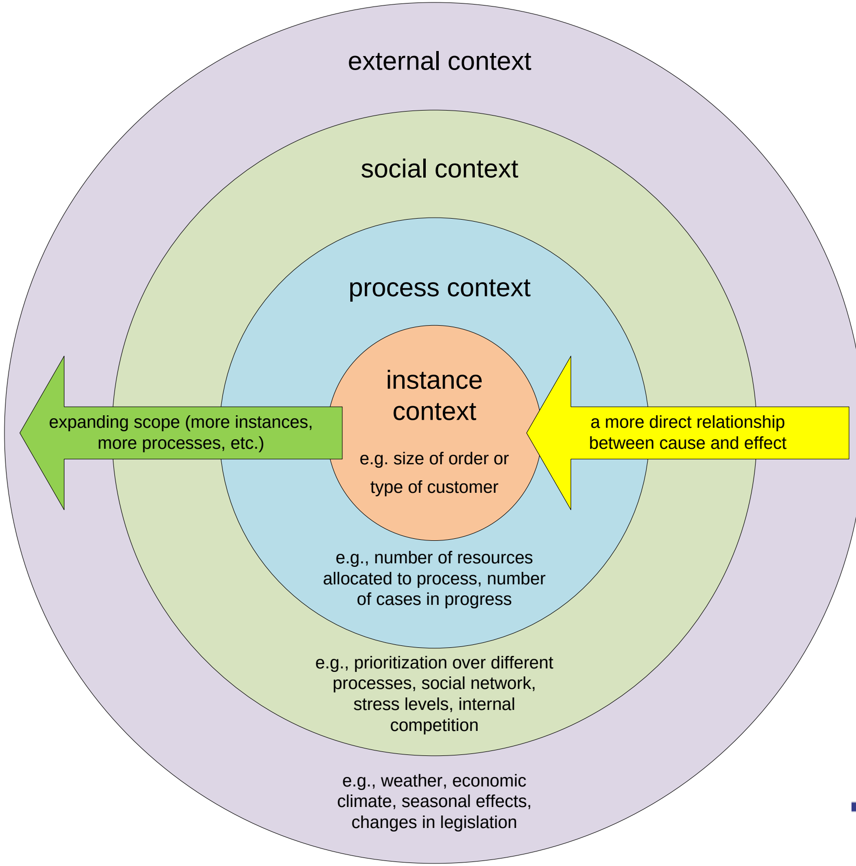


Where do the predictor variables come from?



Predictor variables may also be based on the context of the process instance

- Number of cases running (e.g. skip check if busy).
- Number of resources present.
- Workload of resource.
- Day of the week.
- Weather.



curse of dimensionality

more variables,
more combinations,
data gets sparser
(less instances per
combination),
danger of overfitting,
...


remaining
flow time
service levels
costs
incidents
risks
fraud
"next activity" is
just one of many
possible response
variables !!!

operational support: predictive analytics for processes

Part I: Preliminaries

Chapter 1
Introduction

Chapter 2
Process Modeling and Analysis

Chapter 3
Data Mining

Part II: From Event Logs to Process Models

Chapter 4
Getting the Data

Chapter 5
Process Discovery: An Introduction

Chapter 6
Advanced Process Discovery Techniques

Part III: Beyond Process Discovery

Chapter 7
Conformance Checking

Chapter 8
Mining Additional Perspectives

Chapter 9
Operational Support

Part IV: Putting Process Mining to

Chapter 10
Tool Support

Chapter 11
Analyzing “Lasagna Processes”

Chapter 12
Analyzing “Spaghetti Processes”

Part V: Reflection

Chapter 13
Cartography and Navigation

Chapter 14
Epilogue