



Preprocessing and process mining for business-to- business sales forecasting

Petteri Paananen, Osmo Kauppila & Susanna Pirttikangas

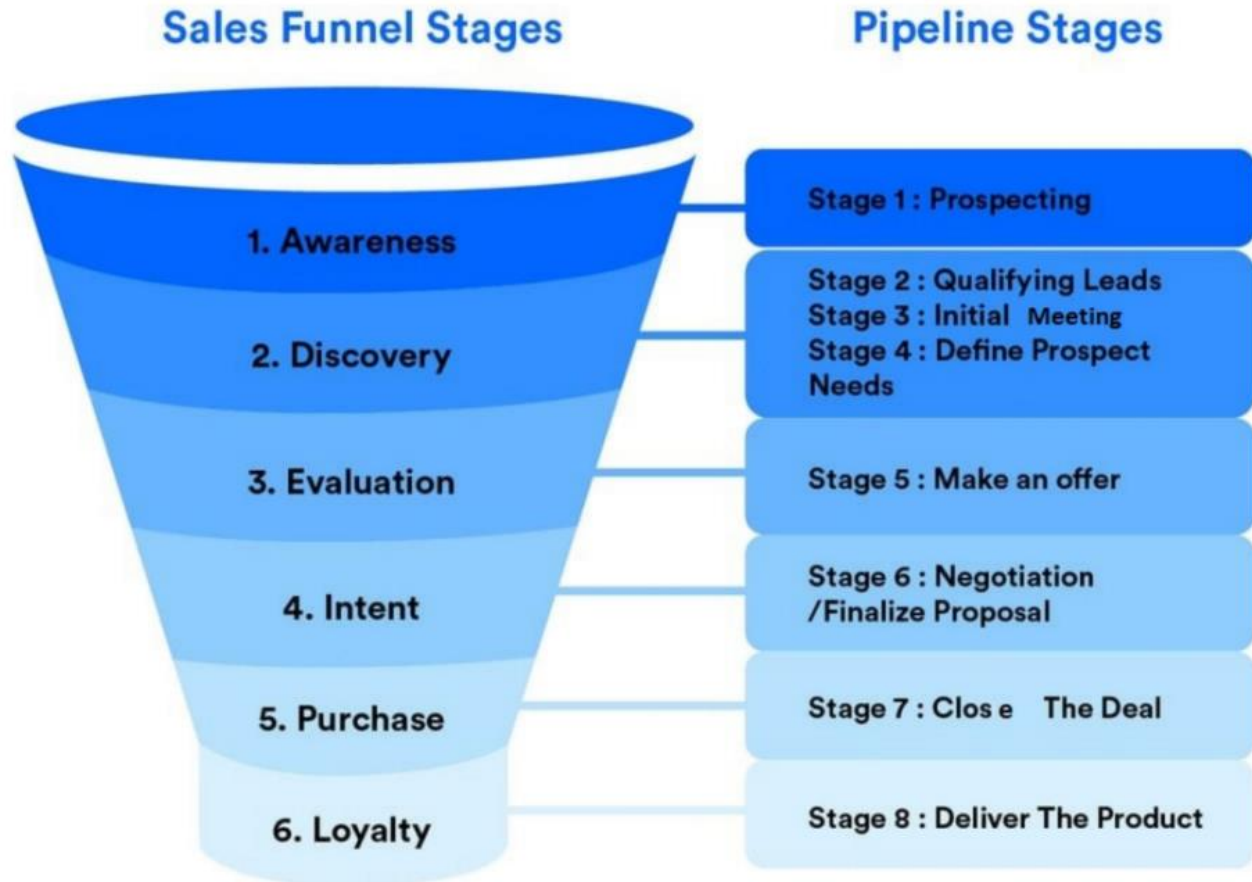
ENBIS 28.6.2022



Forecasting – Problems with Build-to-Order

- Build-to-Order (BTO):
- **Products are tailored for each customer**
- There is no "standard product", historical demand challenging to use
- Production volumes usually low
- Still, forecasts needed in order to manage raw material pricing, capacity and delivery times

Sales Process: Same but Different



– Sales process contain **Stages**.

1. Lead

2. Offer creation / in progress

3. Offered

4. Negotiation

5. Deal / No Deal

– Actual **Stages** vary between companies.

– The process and actual behaviour differ.

Illescas-Manzano, M.D., Vicente López, N., Afonso González, N., Cristofol Rodríguez, C.: Implementation of Chatbot in Online Commerce , and Open Innovation. Journal of Open Innovation: Technology, Market, and Complexity 7(125) (2021). <https://doi.org/https://doi.org/10.3390/joitmc7020125>

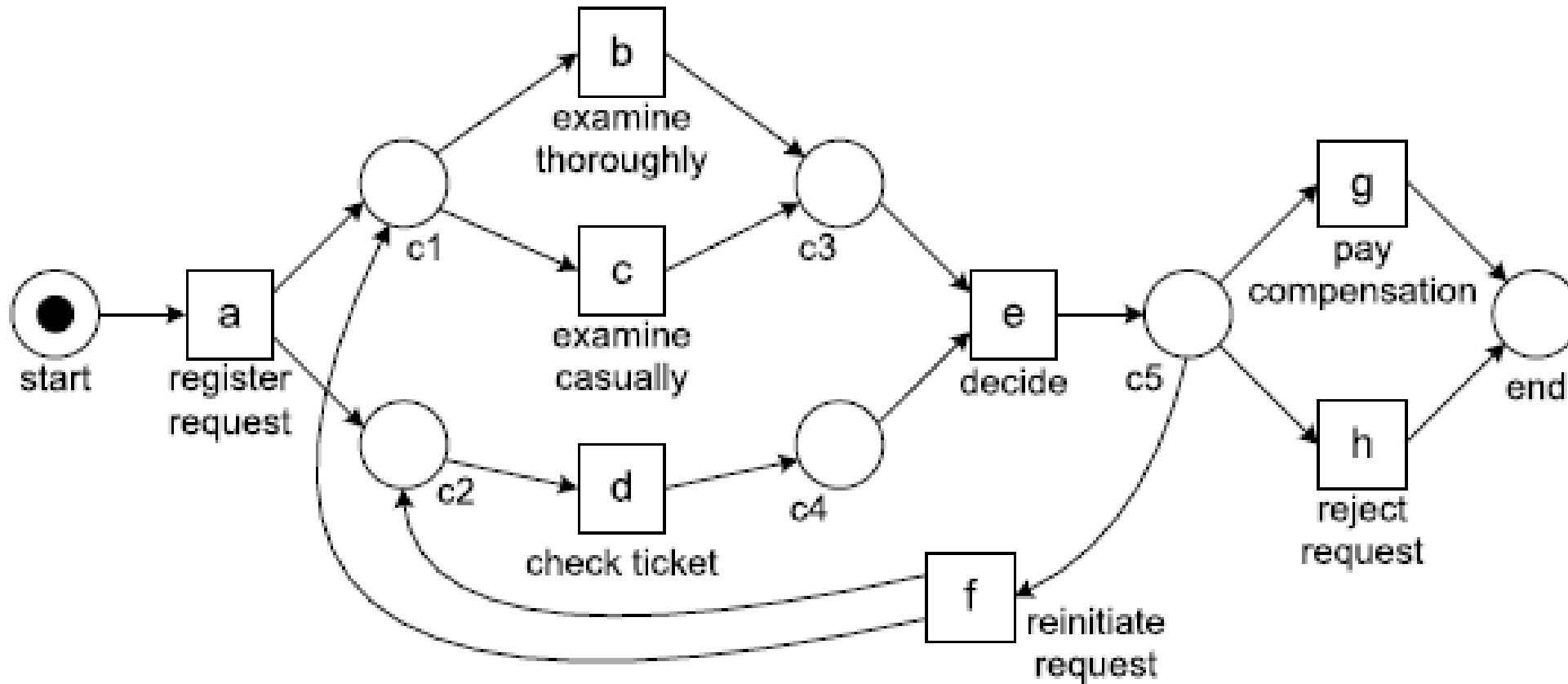


Basic building blocks of a Process log

| Case id | Event id | Properties | | | | |
|---------|----------|------------------|--------------------|----------|------|-----|
| | | Timestamp | Activity | Resource | Cost | ... |
| 1 | 35654423 | 30-12-2010:11.02 | register request | Pete | 50 | ... |
| | 35654424 | 31-12-2010:10.06 | examine thoroughly | Sue | 400 | ... |
| | 35654425 | 05-01-2011:15.12 | check ticket | Mike | 100 | ... |
| | 35654426 | 06-01-2011:11.18 | decide | Sara | 200 | ... |
| | 35654427 | 07-01-2011:14.24 | reject request | Pete | 200 | ... |
| 2 | 35654483 | 30-12-2010:11.32 | register request | Mike | 50 | ... |
| | 35654485 | 30-12-2010:12.12 | check ticket | Mike | 100 | ... |
| | 35654487 | 30-12-2010:14.16 | examine casually | Pete | 400 | ... |
| | 35654488 | 05-01-2011:11.22 | decide | Sara | 200 | ... |
| | 35654489 | 08-01-2011:12.05 | pay compensation | Ellen | 200 | ... |



A process diagram from the log

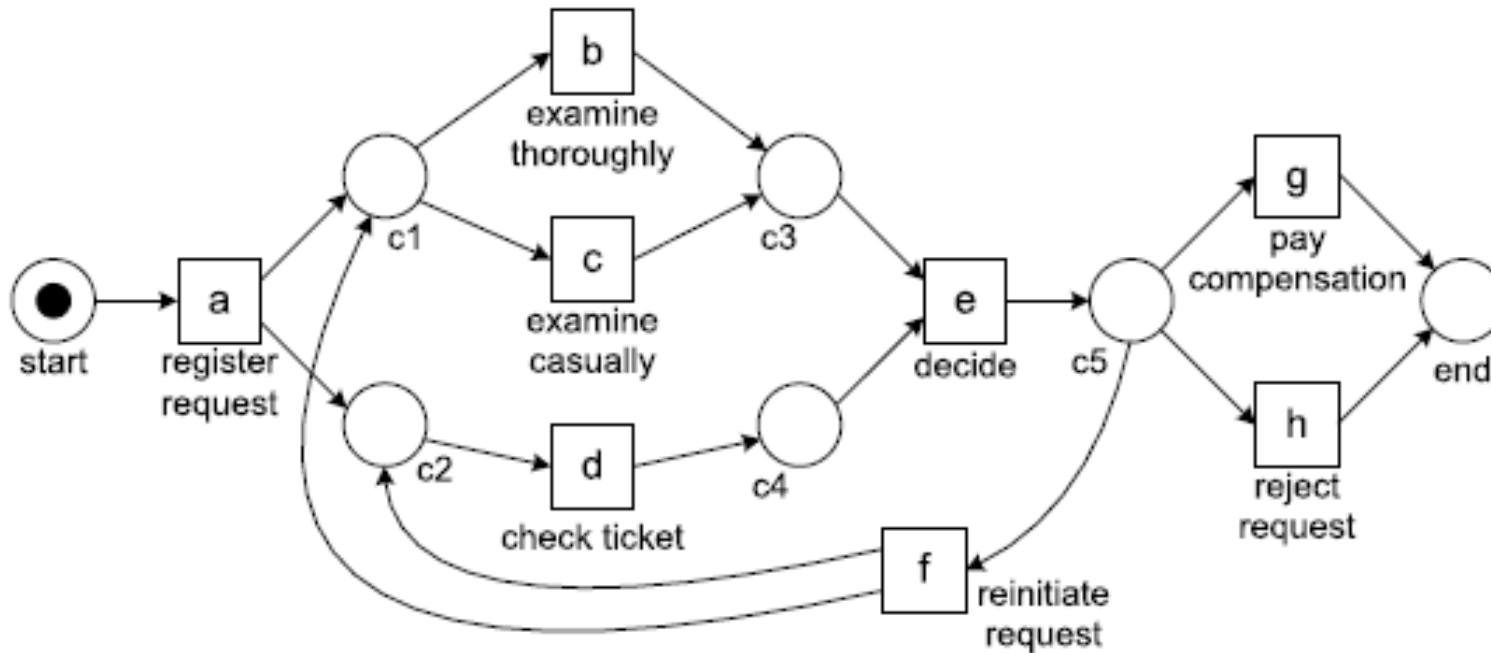


Activity

register request
examine thoroughly
check ticket
decide
reject request
register request
check ticket
examine casually
decide
pay compensation



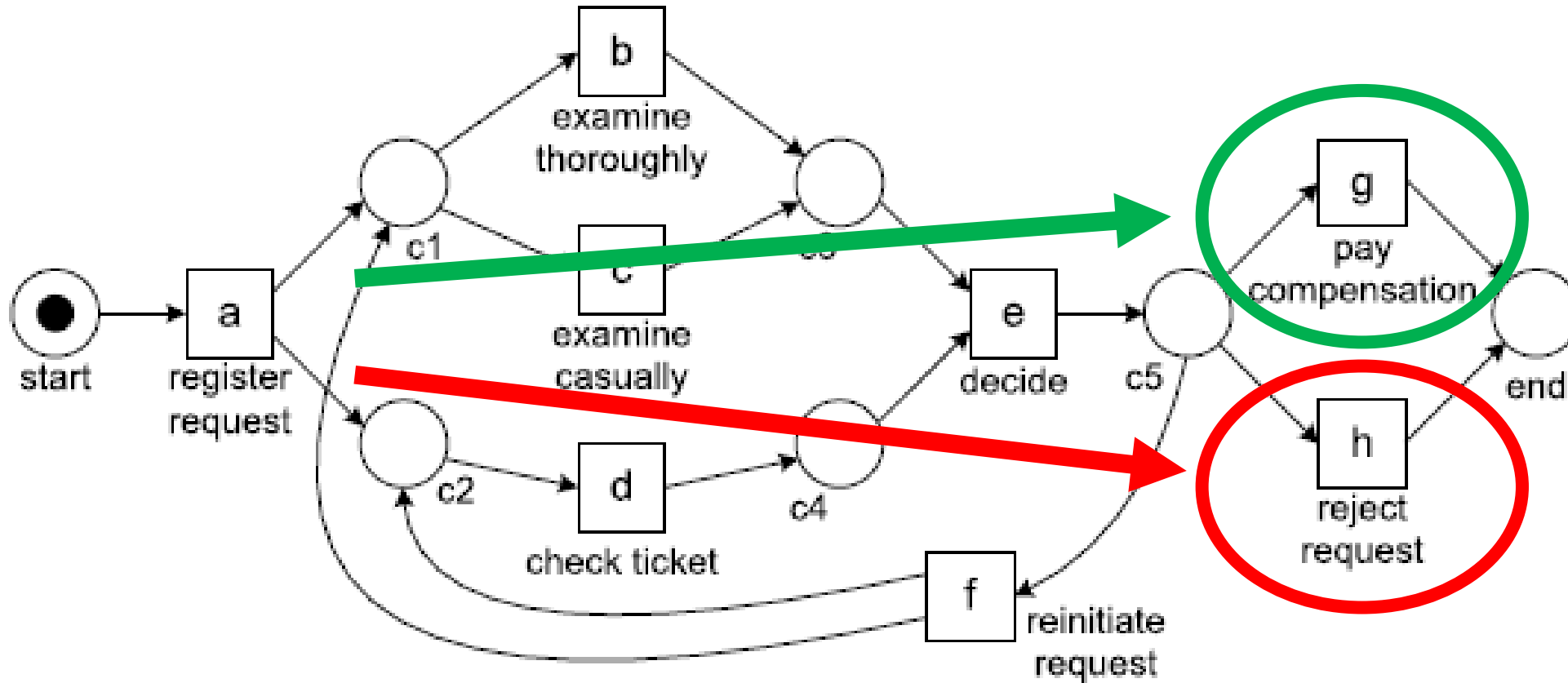
What to Predict? Predictive Process Monitoring



- Numeric
 - Time to next state
 - Time to complete
- Categorical
 - Outcome
- Sequence
 - Next event



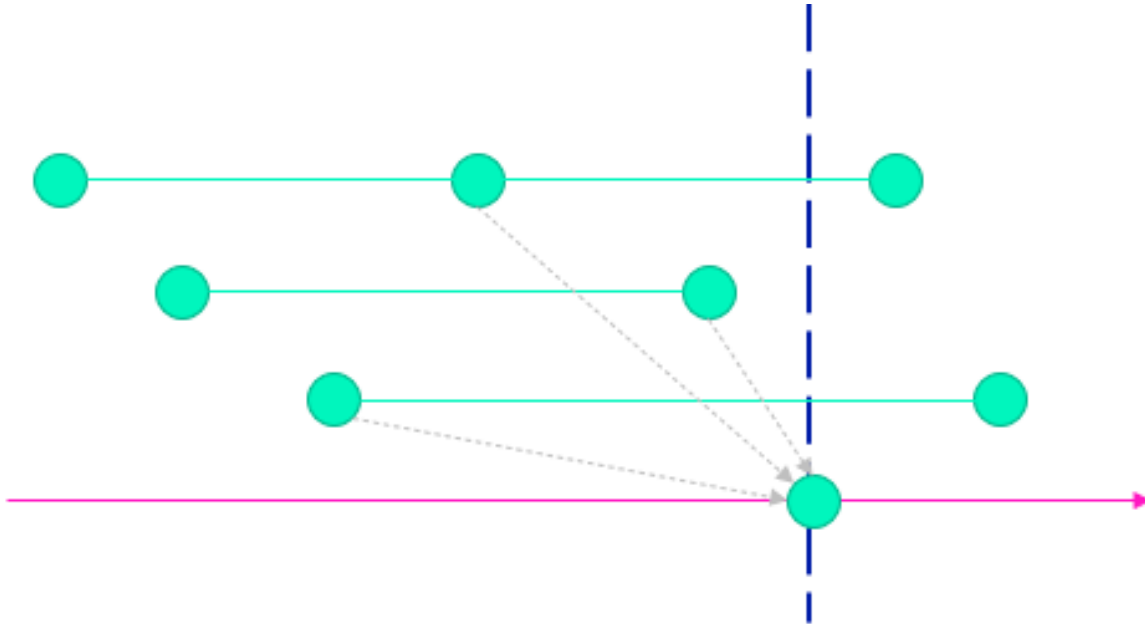
Sales Process Prediction: Outcome





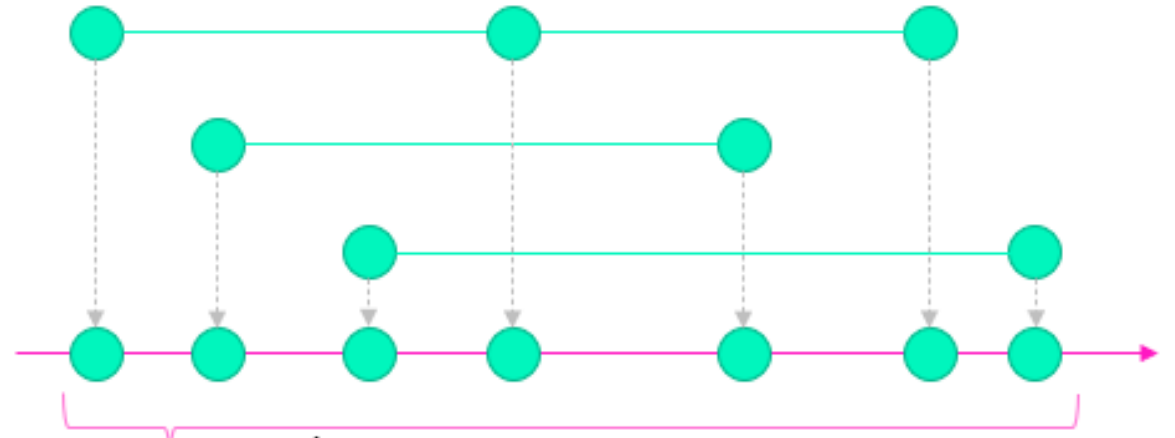
Snapshot or History?

– SNAPSHOT



- Use
 - Labeling attributes at a specific time

– PROCESS HISTORY



- Use
 - Labeling attributes
 - State changes
 - Durations of stages



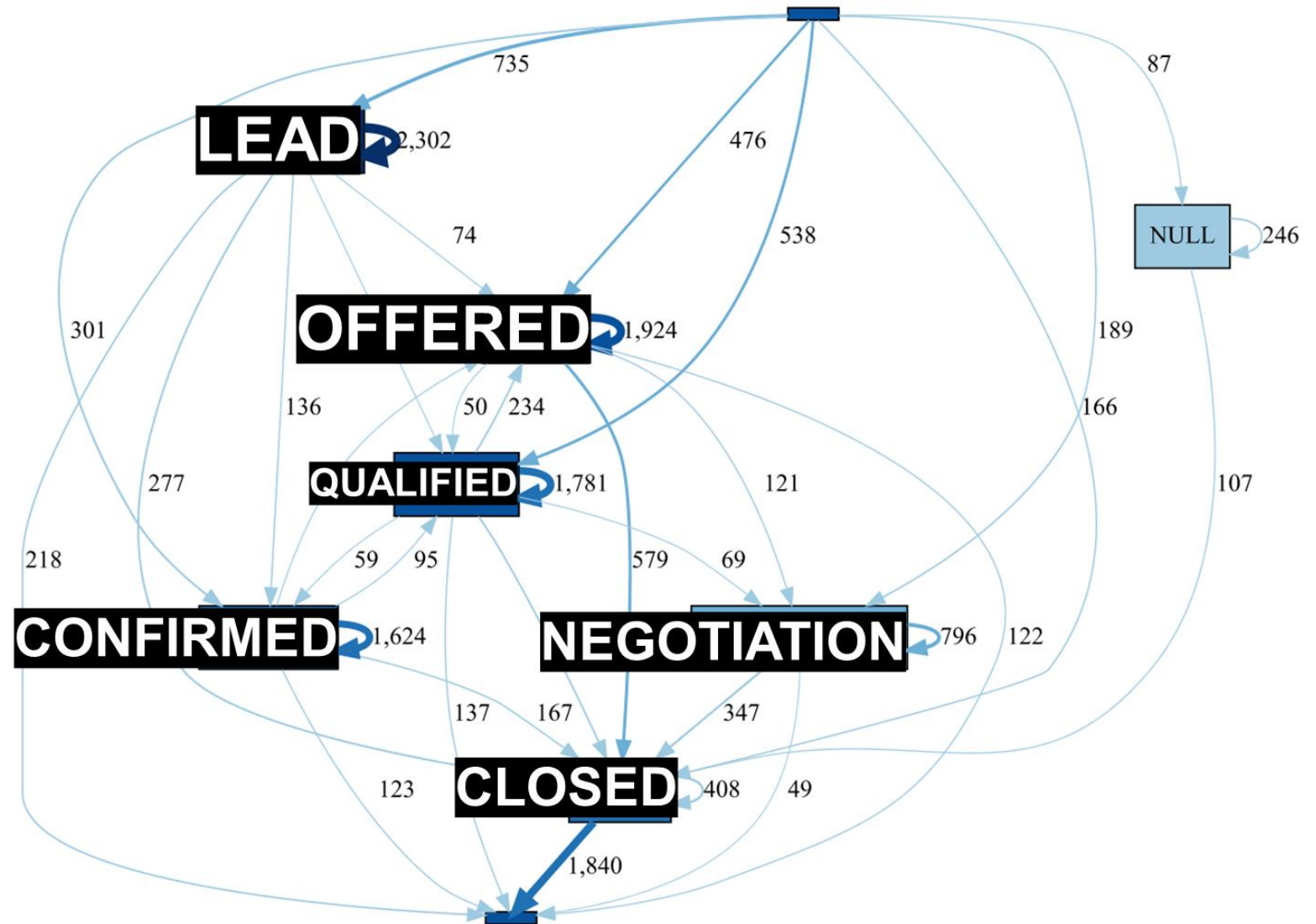
Questions

1. How to predict outcome of B2B sales pipeline where sales items vary highly between customers?
2. How real world event logs behave?
3. Can the prediction be more general? Can we use the same methods (and training) for different logs?



Process model from Dataset

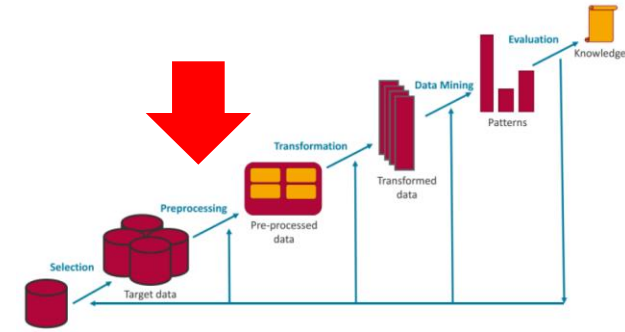
- Real life data from a CRM system
- From a company that sells development projects
- 10 years' worth of data
- 2511 cases
- 14579 events





Preprocessing

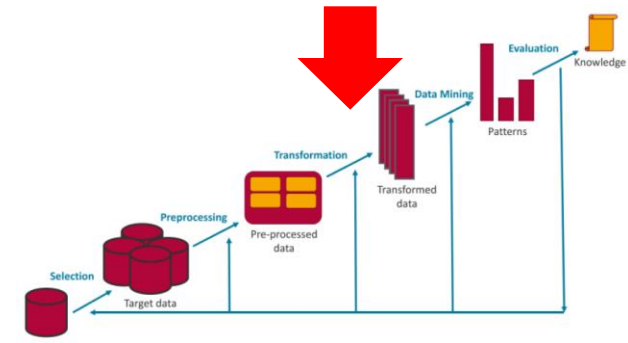
- Remove fields that are not in use
 - Remove cases that do not start with a reasonable state
 - Remove cases that are incomplete
 - Ignore duplicate states – to be tested
 - Impute missing values
 - Label data with win/lose information
- Originally **2511** cases – after preprocessing **1703** cases





Process mining features

- How to encode the information from a process log so that it can be used in prediction algorithm?



Well...



Features can be created in many ways

- Bucketing: Create buckets of first events -> tailor predictions for different buckets
- Code few events together
- Use state changes
- Use frequencies
- Use dynamic features
- Use Hidden Markov Models -> generate probabilities for event sequences that cannot be observed directly
- Use times between events
- Use repeating same events
- ... How to select the right ones?



Further research ideas

- How to optimize the methods to be used for feature creation?
Automatically?
- Transfer learning – can we train with process from Company A and then predict with Company B process? What it takes?
- What architectures we could use? Not just a single Python script?



Summary

- Build-to-Order products forecasting needs special methods
- Process mining and Predictive Process Monitoring can be used
- Real life datasets are messy, a lot of preprocessing needed
- Selecting right set of features (and ways to create features) requires a lot of experimentation



Thank you!

Questions? Please contact us at:

petteri.paananen@oulu.fi

osmo.kauppila@oulu.fi

susanna.pirttikangas@oulu.fi