



Contribution ID: 70

Type: **not specified**

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

*Tuesday, 28 June 2022 17:10 (20 minutes)*

When products are made-to-order, sales forecasts must rely on data from the sales process as a basis for estimation. Process mining can be used for constructing the sales process from event logs, but it also provides characteristics from the process itself that can be utilized in prediction algorithms to create the actual sales predictions. Based on literature, the encoding of process information has a larger impact on the accuracy of the prediction than the actual algorithms used. Most of the studies available on this topic use standard data sets, which provide a safe ground for developing encoding methods and testing different prediction algorithms. It is argued, however, that results from a single data set are not transferable to real world data sets.

As the type of event log affects the prediction results significantly, sales process' outcome prediction should be investigated more thoroughly. All sales processes contain somewhat similar stages such as lead, offer, negotiation and similar attributes, such as a categorization of the customer, sales manager, categories of the product, probability of the positive outcome. In previous work, a plethora of techniques are provided for pre-processing and encoding. To take these findings into practice, they need to be applied for real world data sets in the sales process domain. Which methods work and why? Are there some methods of pre-processing and encoding that can be generalized for all sales processes? This study aims to enrich the understanding of these topics based on one real-life data set.

### Keywords

preprocessing, process mining, sales

**Primary author:** Mr PAANANEN, Petteri (University of Oulu)

**Co-authors:** KAUPPILA, Osmo (University of Oulu); Dr PIRTTIKANGAS, Susanna (University of Oulu)

**Presenter:** KAUPPILA, Osmo (University of Oulu)

**Session Classification:** CONTRIBUTED Reliability 3 + Mining

**Track Classification:** Mining