



**Hochschule für Technik
und Wirtschaft Berlin**

University of Applied Sciences

Project Report

Student Project Business Intelligence – Process Causality Analytics

Field of study: Business Intelligence

Processing period: May 18 2021 - September 17 2021

Fabio Cosimo Andriulo 575311

Timur Burkholz 575308

Maximilian Alexander Reech 575311

Professor: Prof. Dr. Ingo Claßen

Supervisor: Benjamin Aunkofer

Partners (Signavio): Timotheus Kampik, Konstantinos Poullos

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Contents

List of Figures	II
List of Tables	III
1 Introduction	1
1.1 Background	1
1.2 Goal	1
1.3 Structure	1
2 Double Machine Learning	2
2.1 In general	2
2.2 In this case	3
3 Data acquisition	5
3.1 Real Data	5
3.2 Simulation	5
4 Implementation	10
4.1 Simulation	10
4.2 Data Transformation	11
4.3 Double Machine Learning	12
5 Conclusion	13

List of Figures

2.1	Process 1 (Unchanged)	3
2.2	Process 2 (Changed)	3
3.1	BPMN: Process 1 (Unchanged)	7
3.2	BPMN: Process 2 (Changed)	8
4.1	Bildunterschrift	10

List of Tables

4.1	Tabellenunterschrift	11
-----	--------------------------------	----

1 Introduction

This document was created as a report about the project. The project is part of the Master's program in Business Informatics at HTW Berlin. The project has been presented by DATANOMIQ GmbH and was accompanied by Signavio GmbH as a cooperation partner.

1.1 Background

As a part of business intelligence, process mining is a way for modern companies to get information about their processes. It is not uncommon for processes to be further adapted and changed over time. The question always arises to what extent the changes have influenced the measurement results of the process.

1.2 Goal

In order to specify this question further, it should be found out whether there is a causal relationship between the change in the process and the change in the measurement result. In the best case, the goal is to find the specific changes from all the changes that also influence the measurement result.

1.3 Structure

First, the methods used are introduced. This is to clarify afterwards, for which reasons which kind of data was used. Afterwards, the implementation of the software will be discussed in order to explain its peculiarities, limitations and possibilities.

2 Double Machine Learning

2.1 In general

A data-based causal analysis attempts to measure the effect of a change on observable outcome. The measurement of this effect is possible when the changes are observable. This can be implemented by double machine learning. In general, it tries to estimate the effect of the change on the outcome by the changes.¹ In order to recognize the causal effect of a change, all other influences must be comparable.² In addition, it is necessary that the results can arise from the same characteristics, as well as there should be no combination that is found only before the change or after the change.³

By using machine learning, the relationships between the changes and their effects can be represented as machine models. It should be noted that the representation as a model is always only an approximately correct estimate of the real relationship.⁴ Now we can assume that there are two models. One to calculate the result under condition of the change and one to explain the change. By assuming that a change is calculable if the calculation of the first model is different under the different conditions (changed/not changed), it can be assumed that the representation of the change by the second model explains the difference.⁵

¹see Huber 2020, p. 106.

²see [ibid.](#), p. 107.

³see [ibid.](#), p. 109.

⁴see [ibid.](#), p. 111.

⁵see [ibid.](#), pp. 112 sqq.

2.2 In this case

In this case, however, we are talking about process changes. If there is a suitable representation of the process properties in the flow and its results, then double machine learning can also be used for this. In the following, the idea is explained with an example.

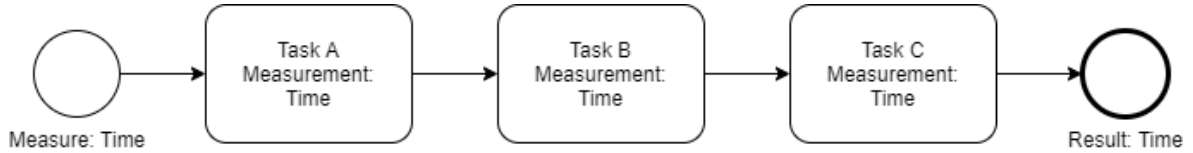


Figure 2.1: Process 1 (Unchanged)

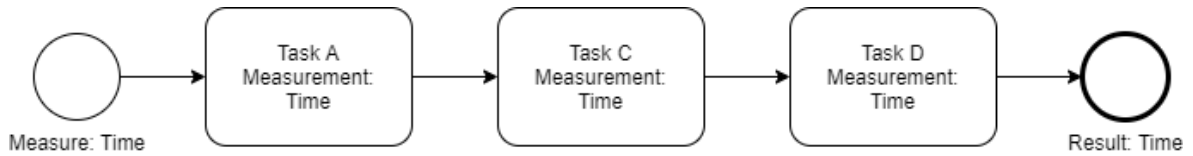


Figure 2.2: Process 2 (Changed)

As can be seen in the figures, Task B was deleted and Task D was introduced in the modified process. As a simple example, we now use time as our measurand. Here, the result of the measurement is the sum of the time used for the tasks. We can therefore set up the following example functions:

p_1/p_2 representing the results, $A..D$ representing the measurements

$$p_1 = A + B + C$$

$$p_2 = A + C + D$$

Now, we can assume that a machine learning model (u) can be trained for the unmodified process as the adornment of the estimate p_1 and as the characteristics $A..C$. Then, we can try to use the model (u) to perform a prediction for our modified process to measure the difference (β) in the outcome.

$$u \approx p_1$$

However, this requires the model to make a prediction (q) for the outcome of the changed process.

$$u \rightarrow p_2 \approx q$$

Now the difference (β) can be calculated.

$$\beta = p_2 - q$$

A second model (m) shows the effect of the change on the difference.

$$m \approx \beta$$

In this case, it is clear because the change in results can only be explained by task D.

$$m \equiv D$$

In more complex examples, it is conceivable that the change in difference may also consist of combination of changed characteristics. In the end, it is only necessary to find out whether the estimation of the difference was successful. A classical regressive measurement would be the mean squared error. The smaller this is (closer to zero) the better the estimation succeeds and thus the presentation as a causal reason.

$$m \rightarrow \beta \equiv \beta_m$$

$$causality = mean((\beta - \beta_m)^2)$$

3 Data acquisition

3.1 Real Data

To implement such an algorithm, you need data to work with. In the best case, real data is available from a company that has just adapted its processes. The difficulty is that there must be at least two data sets from the same process. One must represent the state before the changes and one after the changes. Since no data could be provided during the course of the project, some had to be sought. One place to start is the Business Process Intelligence Challenge, which annually challenges participants with problems in process mining. However, even after finding two data sets, with the same origin and the same process, problems arose. On the one hand, the datasets were in different languages and on the other hand, some of them were poorly described. This is due to the time gap of five years (2012¹, 2017²). Therefore, it was decided to simulate data.

3.2 Simulation

During the project, a simple order-to-cash process was designed in BPMN-format. The order-to-cash process encompasses all steps from when a customer order is placed up until the business is paid (the cash). Those steps include order management and order fulfillment, through to credit management, then invoicing and ultimately payment collection. Since the project does not focus on developing a complex process, a simplified process was used.

A variety of sources were consulted for this purpose and finally the process from the following source was used: Dumas et al. 2018, p. 373.

Other processes from Fahland et al. 2020, p. 251 and Dubois 2017, p. 353 were under consideration, but were discarded as the project progressed.

The BPMN process shows the process from ordering a product to delivery. The example company

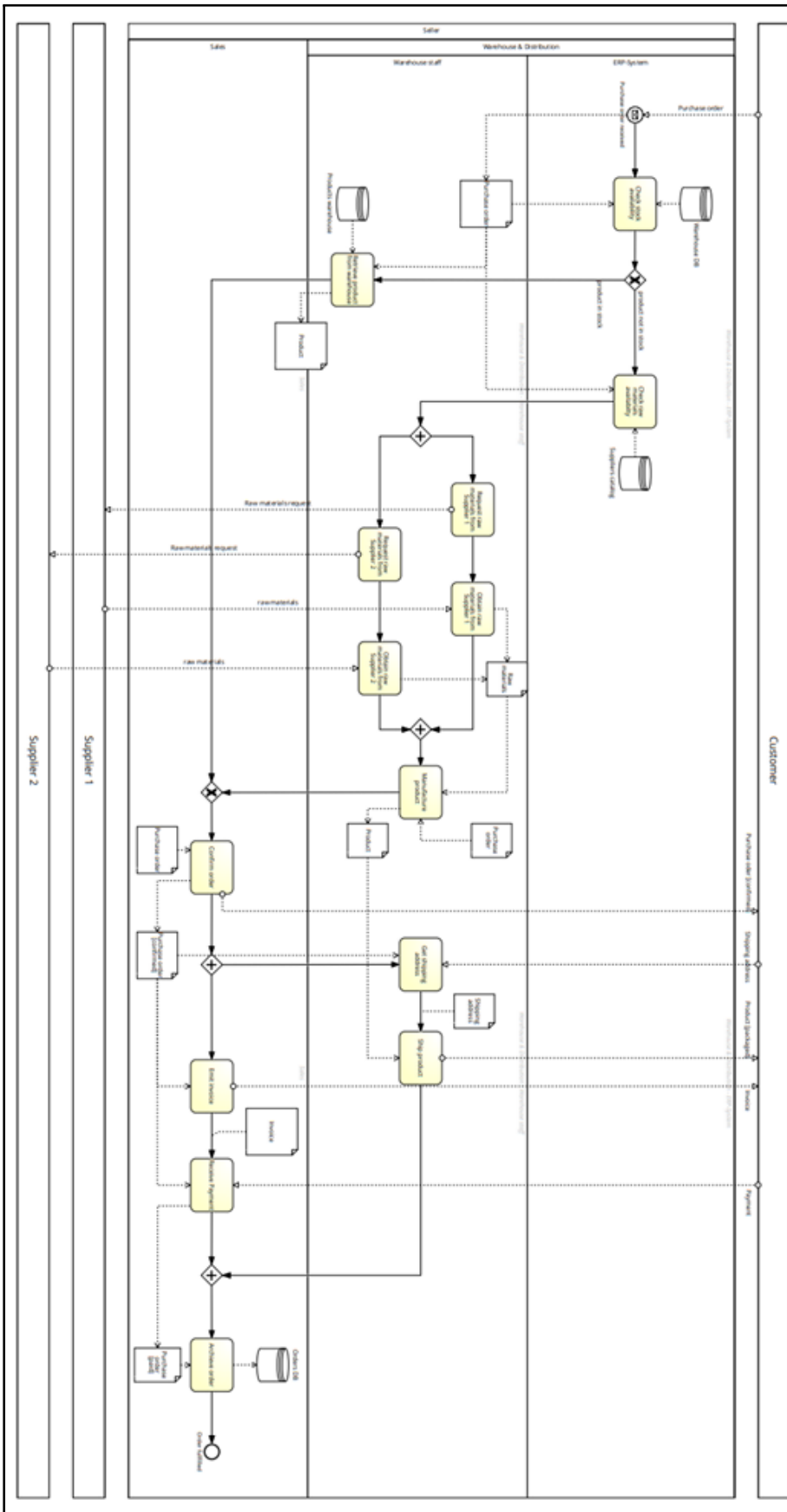
¹BPI 2012 <https://www.win.tue.nl/bpi/doku.php?id=2012:challenge>

²BPI 2017 <https://www.win.tue.nl/bpi/doku.php?id=2017:challenge>

works with two suppliers. The three roles of the ERP system, a warehouse employee and a sales employee are represented by swim lanes. When a customer orders a product, it is checked whether it is already in stock. If it is, the product is ordered from the warehouse and confirmed after the order is received. If the product is not in stock, the raw materials are ordered from the two suppliers and then the product is manufactured. If the product is then available, the product is shipped to the customer's delivery address. At the same time, an invoice is sent to the customer. When the amount is received and the product has been delivered, the order is archived, and the process is finished.

The process is considered the basis for the project. For this purpose, a second similar process will be developed, in which the first process will be modified.

The BPMN process can be found on the next page.





In the previous page (8) you can see the customized process. The markers show the changes.

The following adjustments were applied:

1. A new gateway was integrated. Here it is checked whether the raw materials are available at the suppliers. If this is not the case, a new end is reached and the customer is informed that the product is not available.
2. Orders for raw materials have now been automated by the ERP system. A warehouse employee no longer must do this manually. For this purpose, the suppliers' databases were connected and automatic processes were started.
3. The customer's delivery address is now entered directly by the sales staff, who also confirm the order. In addition, this changed the sequence and parallelism of the process steps.

4 Implementation

4.1 Simulation



Figure 4.1: Bildunterschrift

Erstellt in Eigenarbeit in Anlehnung an see Unternehmen Sonstige GmbH [2020](#)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. (siehe Abbildung 4.1)

4.2 Data Transformation

Inhalt	Jemand musste Josef K. verleumdet haben, denn ohne dass er etwas Böses getan hätte, wurde er eines Morgens verhaftet. “Wie ein Hund!” sagte er, es war, als sollte die Scham ihn überleben.
Inhalt	Jemand musste Josef K. verleumdet haben, denn ohne dass er etwas Böses getan hätte, wurde er eines Morgens verhaftet. “Wie ein Hund!” sagte er, es war, als sollte die Scham ihn überleben.
Inhalt	Jemand musste Josef K. verleumdet haben, denn ohne dass er etwas Böses getan hätte, wurde er eines Morgens verhaftet. “Wie ein Hund!” sagte er, es war, als sollte die Scham ihn überleben.
Inhalt	Jemand musste Josef K. verleumdet haben, denn ohne dass er etwas Böses getan hätte, wurde er eines Morgens verhaftet. “Wie ein Hund!” sagte er, es war, als sollte die Scham ihn überleben.
Inhalt	Jemand musste Josef K. verleumdet haben, denn ohne dass er etwas Böses getan hätte, wurde er eines Morgens verhaftet. “Wie ein Hund!” sagte er, es war, als sollte die Scham ihn überleben.

Table 4.1: Tabellenunterschrift

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc,

molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. (siehe Tabelle 4.1)

4.3 Double Machine Learning

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.¹

¹see Literatur 2020, p. 1.

5 Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Bibliography

- Dubois Eric; Pohl, Klaus (2017). *Advanced Information Systems Engineering*. Springer International Publishing.
- Dumas, Marlon et al. (2018). *Fundamentals of Business Process Management*. Springer Berlin Heidelberg.
- Fahland, Dirk et al. (2020). *Business Process Management Forum*. Springer International Publishing.
- Huber, Martin (2020). *Kausalanalyse mit maschinellem Lernen*. Springer Fachmedien Wiesbaden GmbH.
- Literatur, Vorname (2020). *Titel Literatur*. Springer Vieweg. ISBN: 123-4-567-89101-1.
- Unternehmen Sonstige GmbH (2020). *Titel Sonstige*. URL: <https://www.google.com> (visited on 10/14/2020).