Design and deployment of an analytic artefact–investigating mechanisms for integrating analytics and manufacturing execution system

investigating mechanisms for integrating analytics and manufacturing execution system

Asmussen, Claus Boye; Jørgensen, Steffen Lundgaard; Møller, Charles

Published in:
Enterprise Information Systems

DOI (link to publication from Publisher):
10.1080/17517575.2021.1905881

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Asmussen, C. B., Jørgensen, S. L., & Møller, C. (2022). Design and deployment of an analytic artefact–investigating mechanisms for integrating analytics and manufacturing execution system: investigating mechanisms for integrating analytics and manufacturing execution system. *Enterprise Information Systems*, *16*(10-11), 1587-1616. Advance online publication. https://doi.org/10.1080/17517575.2021.1905881

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

Design and deployment of an analytic artifact – investigating mechanisms for integrating analytics and manufacturing execution system

*Claus Boye Asmussen (cba@mp.aau.dk)**

*Steffen Lundgaard Jørgensen (Steffenmail@gmail.com)**

*Charles Møller (charles@mp.aau.dk)**

**Department of Materials and Production, Center for Industrial Production, Aalborg University, Fibigerstræde 16, 9220 Aalborg Øst, Denmark*

Corresponding Author: Claus Boye Asmussen (cba@mp.aau.dk)

Word count: 10695

Design and deployment of an analytic artifact – investigating mechanisms for integrating analytics and manufacturing execution system

Abstract:

Integrating Enterprise Information Systems (EIS) into companies have been researched over the last decades and is considered well understood. However, the integration of analytics and EIS is lacking the same research interest, which is increasingly becoming important due to the need for analysing enterprise data in order to understand and manage the complexities of the supply chain and in particular manufacturing. Correspondingly, this paper investigates the mechanisms for integrating an analytic artefact with Manufacturing Execution System (MES), by constructing and deploying the artefact within two manufacturing sites. The findings indicate that analytic artefacts, unlike EIS or IT integrations, can be constructed based on a business problem and do not have to address specific factors such as IT-systems, process, people, or tasks. As a consequence, an analytic artefact can be built centrally and deployed across heterogeneous manufacturing sites, if they share a common data model.

Keywords: analytics, SCM, EIS, MES, action design research, machine learning

1 Introduction

Managing a company in today's environment is extremely difficult, as supply chains are becoming more complex, as there is a need to reduce cost, increase quality, and flexibility to become and stay competitive. Much of the complexity stems from the ever increasing information and data flows introduced to the supply chain (Bose 2009; Beer 2018). Especially manufacturing companies have the potential to generate massive amounts of data and consequently possess an untapped potential to transform manufacturing data into actions to create a competitive advantage. Analytics is used to address this complexity, by analysing data and identifying, so far unknown relationships.

The data of a supply chain company is stored and managed in Enterprise Information Systems (EIS), which for analytical purposes, must be integrated with specialised analytics software, to deal with the complexity of the supply chain enterprise data, and to be able to do advanced analytics e.g. advanced planning, what-if analysis, and correlations analysis of complex systems, which is not widely adapted in industry (Link and Back 2015; Ranjan, Jha, and Pal 2016; Koh, Gunasekaran, and Goodman 2011; Benaissa and Benabdelhafid 2005; Addo-Tenkorang, Helo, and Kantola 2017; Akyuz and Rehan 2009; Gupta and Jones 2014).

Further, there is a rising research interest within the theme of 'factories of the future', which includes analytics as a sub-part, where legacy IT-systems need to be integrated with service-oriented streams (Pei Breivold 2020). A part of the 'factories of the future' is to make manufacturing smarter by e.g. making use of (big data) analytics (Dai et al. 2020) and enable the use of cognitive computing (Kumar Sangaiah et al. 2020). However, the application of analytics in enterprises appears fragmented and lacks clarity (Khanra, Dhir, and Mäntymäki 2020). Additionally, through a literature review Barbosa et al. (2017) finds that research within the areas of supply chain and analytics is a premature research area, where most research have been within the areas of demand planning and order fulfilment. They find that a research issue in dealing with the integration between supply chain IT systems such as EIS, and analytical artefacts and recommend research within this scope.

Thus, the aim of this paper is to investigate the mechanisms and gain learning and findings of integrating an analytic artefact with an EIS. This paper addresses the issue by constructing and deploying an analytical artefact with Manufacturing Execution Systems (MES) in an Action Design Research (ADR) methodology. The findings and learnings are identified by constructing and deploying an instance of an analytic artefact in two manufacturing sites. The construction and deployment of the analytic artefact enables the research to address two aspects: one, to ensure the practical relevance of the research and two, to prescriptively generate knowledge and learnings of the mechanisms needed for integrating analytics and MES. It is expected that the findings from the construction and deployment of the artefact within the two manufacturing sites, can be used to understand the mechanism of integrating analytics and MES for manufacturing sites in general and integrating analytical artefacts with other EIS. The analytic artefact is constructed as an IT artefact. The analytic artefact is designed to incorporate the advanced levels of analytics, i.e. predictive and prescriptive analytics, and solve a business problem defined by the workers at the first manufacturing site.

The structure of the paper is as follows. Section 2 presents the methodology, Section 3 presents a literature review and section 4 presents the construction and deployment of the artefact in the first

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

manufacturing site. Section 5 presents the evaluation of the findings from the construction of the artefact by deploying the artefact in the second manufacturing site. The paper is concluded by a discussion in section 6 and a conclusion in section 7.

2 Methodology

Identifying the mechanisms for integrating analytics and MES requires the research to not only have the expected academic rigor, i.e. ensure the reliability and validity of the research by reviewing the knowledge base, but also do empirical research, as the research problems stems from a practical problem. Further, the integration of analytics and MES is not widely used in the industry, which makes it difficult to apply the classical explanatory sciences. Instead, the use of prescriptive methodologies, enable the researchers to create solutions to problems, obtain learnings, and provide recommendation for solutions to the problem.

ADR is chosen as a methodology as it provides the necessary guidance in developing and evaluating an artefact and recognising the importance of the organisational context the research is conducted within (Sein et al. 2011).Constructing an artefact on a case company, can provide valuable design knowledge about specific mechanisms that enable a successful analytics integration including not only the technical challenges, but also the organisational context. The design of an artefact changes during the construction and consequently evaluating an artefact must be done continuously and ends once the value and utility of the artefact is satisfactory. The value of ADR lies not in the artefact itself, but in reflection and learning. The evaluation is based on the technical integration between the analytical artefact and MES and not on the user's evaluation of the artefact.

The paper will present the findings in the following way. Section three will provide the academic foundation for which the artefact is constructed upon. The section provides an overview of the analytic process, change in the analytic paradigm, and the application of analytics and MES. In section 4, the artefact is constructed based on the literature review and the context of the manufacturing site. The construction of the artefact is based on the CRISP-DM framework. The section is concluded by summarising the findings and learnings about what mechanisms is needed to successfully integrate an analytic artefact with MES. The aim of section 5 is to verify the findings in another environment, by deploying the artefact in another manufacturing site producing different products, with different people, task, and IT-systems. Section 6 discuss the impact of the findings of the mechanism of integrating analytics and MES, as well as the impact on integrating analytic artefacts with EIS, in regards to both an academic and managerial audience. The final section summarised the overall findings and learnings in section 7.

3 Literature review

Identifying the mechanisms of integrating analytics solution with MES, requires an understanding of the analytics process, as well as understanding how analytics and MES is connected. This paper focuses only on the integration of analytics with MES and thus does not cover the general aspects of 'factories of the future' or concepts as smart manufacturing. However, other efforts have been made in dealing with and taking advantage of the increasing data flow. Recent examples are, Zhang and Ming (2020) introduces a reference subsystem for managing multi-process, multi-intersection, and multi-operators with the aim of supporting personalised products and services, and Leng, Jiang, Liu, and Wang (2020) proposes an approach for self-organising manufacturing processes for mass

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

individualisation. While these approaches and concepts are relevant in making better use of the data flows for companies, they do not provide a better understanding of the mechanisms, i.e. behaviours, responses, and processes triggered by intervention of integrating analytics with MES, and as a result the outcome of an intervention is unknown (Jonsson and Holmström 2016). Thus, will the remaining of this section describe the analytical process, as well as how analytics and MES is connected.

Analytic is in this article defined as 'The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. The analytics may be input for human decisions or may drive fully automated decisions. Analytics are a sub-set of [...] business intelligence.' (Davenport and Harris 2007).

The analytics process includes everything from gathering, processing, manipulating, modelling, to visualising data, and is seen as a group of approaches, procedures and tools (Trkman et al. 2010). Analytics can provide solutions to business problems either by descriptive, diagnostic, predictive, or prescriptive methods (Mortenson, Doherty, and Robinson 2014; Herden 2017; Hindle and Vidgen 2018; Nguyen et al. 2017; Porter and Heppelmann 2015; Demirkan and Delen 2013). Most business have or are using descriptive and diagnostic analytics, where predictive and prescriptive analytics become more complex and must be meet with exploratory and experimental iterative approaches, changing the design and course of actions based on insights gained (Herden 2019; Bose 2009; Viaene and Van Den Bunder 2011; Marchand and Peppard 2013; Larson and Chang 2016; Carillo 2017). The analytic approach is notably different than IT approaches, for which analytics projects usually are conducting within (Viaene and Van Den Bunder 2011; Marchand and Peppard 2013).

The process usually starts by identifying and understanding a business problem, by ideally interviewing the end users, which ensures fit to business processes and user's needs, is more likely to be operationalised, and shows better performance (Herden 2019). The next steps include collecting, preparing, analysing the relevant data to finally create a model that solves the business problem (Herden 2019; Provost and Fawcett 2013; Leventhal 2015; Janssen, van der Voort, and Wahyudi 2017; Shearer 2000). The model is then deployed and maintained to continuously support the users of the model (Larson and Chang 2016). The process of constructing analytical artefacts is usually exploratory and thus outcomes of the process is unknown. However, the analytical process can be structured by e.g. use of the CRISP-DM framework (Herden 2019; Kridel and Dolk 2013; Subramaniyan et al. 2018; Nagorny et al. 2017; Shearer 2000).

Some authors argue that the analytics paradigm is shifting away from relying solely on statistical modelling to discover relationships based on correlations from large and feature-rich data (Delen and Zolbanin 2018; Breiman 2001). The new paradigm therefore contradicts the notion that correlation is not causation and one should be careful to draw conclusion, as the correlations could be a coincidence (Delen and Zolbanin 2018). This shift in paradigm is influenced by rise of big data analytics (BDA) where bigger dataset are available, as well as more processing power able to analyse both structured and unstructured data (Barbosa et al. 2017; Gandomi and Haider 2015; Nguyen et al. 2017; Kache and Seuring 2017)

3.1 Analytics and MES

With the recent focus on BDA, the MES system have been identified as a competitive enabler, as it can store data from heterogeneous systems, enabling real-time data analysis via internet of things, advanced analytics and better information sharing for e.g. flexible manufacturing (Cottyn et al. 2011; Weihrauch, Schindler, and Sihn 2018; Jeon et al. 2017; Mansour, Millet, and Botta-Genoulaz 2018; Cupek et al. 2018; De Ugarte, Artiba, and Pellerin 2009). The data gathered in the MES system can be used to make detailed planning and execution on a per production unit basis, as well as create higher degrees of traceability, and is seen as an enabler of predictive and prescriptive analytics. MES is identified as being essential for future supply chain analytics (De Ugarte, Artiba, and Pellerin 2009; Koupaei, Mohammadi, and Naderi 2016; Wang et al. 2018). The current use of analytics in companies is mainly descriptive and diagnostic based on aggregated data (Appelbaum et al. 2017; Hahn and Packowski 2015; Ishaya and Folarin 2012).

A challenge in implementing analytics is that there are no generic analytic solutions available and analytic solutions must be constructed to fit the needs of the specific context. There is therefore a need to create analytic solutions or frameworks to ensure a fit for a given context. To successfully implement analytics in an EIS environment requires IT, supply chain, and managerial knowledge (Asmussen and Møller 2020; Herden 2019).

One of the major challenges in integrating analytics is dealing with system and data heterogeneity. Different IT-systems will often have different data models which leads to errors (Mansour, Millet, and Botta-Genoulaz 2018; Saberi, Hussain, and Chang 2017; Cupek et al. 2018). This is becoming an even greater issue as both structured and non-structured data, along with real-time data is introduced (Saberi, Hussain, and Chang 2017). Several researchers have presented frameworks and models addressing the issue of capturing, managing, and storing real time data for manufacturing companies (Weihrauch, Schindler, and Sihn 2018; Cottyn et al. 2011; Jeon et al. 2017; Cupek et al. 2018; Mansour, Millet, and Botta-Genoulaz 2018; Jiang et al. 2007), however the research have not been empirically proven. Omar, Minoufekr, and Plapper (2019) found prerequisites for enabling analytics in manufacturing data, where data must be standardised, aggregated and stored. These challenges become particular difficult, and highlight an issue with scalability, once the analytics methods are to be used across manufacturing sites, due to the heterogenic data models, systems, and processes. Further, MES is in the industry mostly used reactively, which can be an issue in dealing with the unforeseen events (De Ugarte, Artiba, and Pellerin 2009)

4 Construction of an artefact

The goal of this section is to create an analytic artefact and, in the process, gain valuable design knowledge. The artefact is constructed based on the manufacturing data from a single manufacturing site, where the aim of the artefact is to transform that data into valuable information and actions. The construction of the analytics artefact is based upon the CRISP-DM framework.

As the use of descriptive and diagnostic analytics is prevalent in the industry, and therefore lacks a relevance for both industry and research, entails that the analytics artefact must either be predictive or prescriptive. Further, the analytic artefact should be constructed with the end-user in mind, ideally enabling the end-user to facilitate execution and use of the analytics artefact themselves (Herden 2019).

The analytics artefact must integrate MES, who stores and manages manufacturing data, with tools that has analytic capabilities. The most used analytics tools in industry are the programming languages R and Python, which will be used in constructing the artefact. To enable the use of analytics for manufacturing data requires the data to be standardised and aggregated (Omar, Minoufekr, and Plapper 2019), where the standardisation of manufacturing data in the industry often follows the ISA-88/95 standards. However, it is unclear how these standards can be used for analytics, as they are designed for control and execution purposes (Scholten 2007). Further, an aim is to create a cross-manufacturing site scalable analytic artefact, as researchers have identified scalability as important for value creation, which is thought difficult due to heterogenous data models, IT-systems, and processes. This section creates an analytic artefact addressing these issues, in a manufacturing site, where the academic value lies in the design knowledge obtained during its construction and deployment.

4.1 Case company presentation

The selected case company is an international dairy cooperative who produces a range of products within the product categories of cheese, milk, yogurt, butter, and milk-powder. The company has more than sixty manufacturing sites, also called dairies, with a wide variety in IT systems, data models, processes, people, and tasks. The company has ambitions of using advanced analytics, i.e. predictive and prescriptive analytics, to increase its manufacturing performance, but is having difficulties in generating value. The company is facing many of the same difficulties of managing heterogenous data and IT-systems, where their current analytics approach is based upon having experts create a purpose-built statistical model for each business problem at each manufacturing site, which are proven to be expensive to construct. The result of the approach is that very few manufacturing sites, utilize any form of advanced analytics. As such, the case company is facing several of the identified challenges of integrating analytics with MES and thus is deemed to be a suitable case to study.

The selected manufacturing site, for which the artefact is constructed, is medium sized with about 150 employees and produces blue cheeses, which is classified as a batch type production. The manufacturing processes of the production site is mapped in Figure 1, where the mapping is based on the area level of the ISA-88 standard. Further, the production unit for each production area is presented, as well as the relationship to the next production area. The relationship between production units is important for the batch production, as the production units change a lot from the intake of milk, conversion to cheese curd, maturation of the cheeses, to palletizing the cheeses in the warehouse.

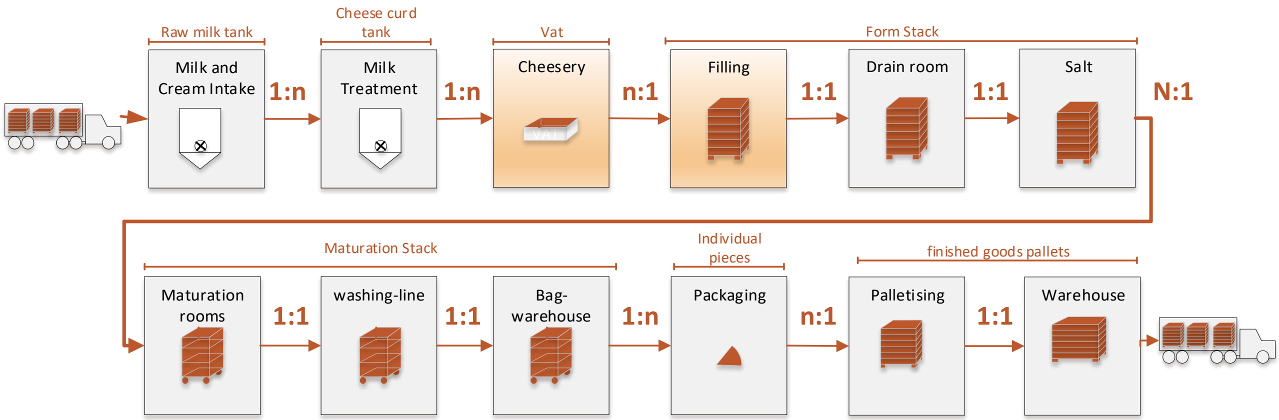


Figure 1 - Overview of the manufacturing site

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

4.2 Constructing an analytics solution for MES

The construction of the analytics artefact follows the CRISP-DM framework, as depicted in Figure 2. The framework is initiated by creating an understanding of a business problem and data, followed by data preparation, modelling, evaluation, and deployment. This section presents the construction of the analytic artefact in the same order, concluding with a section summarising the findings in regards to the identified issues in the literature review. Naturally, the process of constructing the analytic artefact is created in many iterative steps, however only the final solution is presented.

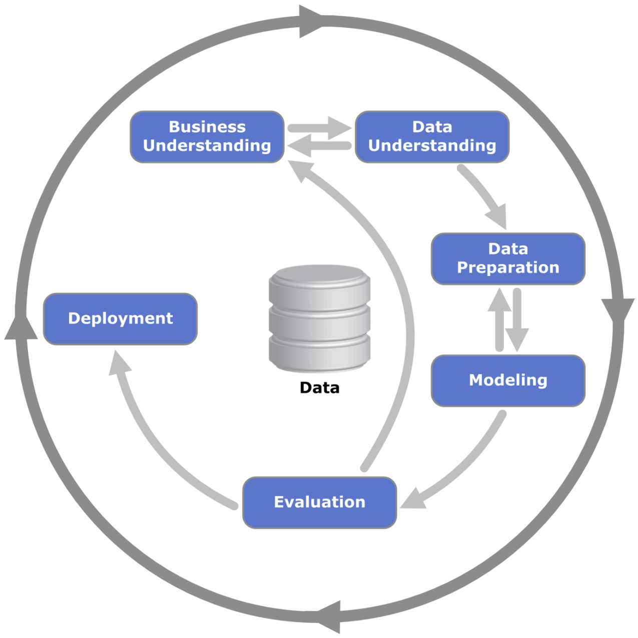


Figure 2 - CRISP-DM framework (Shearer 2000)

4.2.1 Business Understanding

The process is initiated by interviewing key personnel at the manufacturing site, as recommended by (Herden 2019), where specifically the head of production and quality engineer was interviewed. During the interview it became clear that there is a need to create an understanding of the correlations between the different manufacturing processes. Cheese production is essentially producing batches of organic matter, that responds differently to an input depending on the condition of the batch. As an example, measuring a temperature of 30 degrees for a cheese curd means nothing without understanding, e.g. how much culture has been added and the current total production time of the cheese curd batch. As a consequence, it is impossible to create general rules for how to produce cheeses. Further, the complexity of the production of cheeses is so high, even expert dairy engineers cannot fully understand the correlations between the cheese manufacturing processes. Based on the interview and literature review, three aspects are found that an analytics artefact must address, which are:

- Visualise the correlation between production processes on a per production unit level (Explainable AI (xAI))
- Construct a predictive model to answer what-if questions on a per production unit level (Predictive model)

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

- Construct a recommendation model for recommending optimal production values on a per production unit level (Prescriptive model).

The quality engineer specifically mentions the importance of understanding the impact of the production processes on the 24-hour pH value in the filling area for each batch, as too high or low pH values will lead to serious quality issues. The analytic artefact will therefore be addressing the mentioned aspects in regards to the 24-hour pH value.

4.2.2 Data Understanding

To gain an understanding of the data, the quality engineer and a local IT person, identified the relevant data for the analytic artefact. All of the data were structured manufacturing data, based on ISA-88, stored in excel spreadsheets or databases. Each production area has an individual database storing the areas manufacturing data, where some of the areas has excel spreadsheets which are manually updated. The execution and control data for the production is stored in a production database, which contains a batch id, which can be used to join the execution and control data with the areas respective manufacturing data. However, it is not possible in the current setup to identify manufacturing data for the same batch across production areas, as a traceability id is not maintained. Not all excel spreadsheets contain a batch id and as a result cannot be joined to a specific production batch. Only a selection of batches has manufacturing data, as data is recorded by sampling. As a consequence, batches with full data from milk intake until warehousing accounts for 6.5% of all batches. The manufacturing data for each production area is recorded on the areas respective production unit, where e.g. the data in the cheesery is recorded in vat's, a container containing cheese curd, and the data in the filling area is recorded in form stacks, which is several vat's stacked on each other. An overview of the different production unit for the production is depicted in Figure 1.

The pH values relevant to the business problem is stored in the filling area database. Based on the business problem and data understanding it was found that excel spreadsheets should be excluded and limit the selection of databases to the cheesery and filling area databases. Further, the selected data is limited to data recorded in 2018 and for their highest selling cheese which accounts for more than 80 percent of production volume. The reason for the limitations is that most of the manufacturing data is generated within the cheesery area and should be sufficient for a first iteration analytical artefact. Further, both databases have a high rate of sampling per production unit, which ensures a higher percentage of batches with data from both the cheesery and filling area, in this case the percentage of full data on batches is roughly 30%. Though the quality engineer noted she would like to incorporate milk intake data, such as protein and fat content in future iterations.

Based on the business and data understanding two main challenges are identified. One, traceability is needed on a production unit level to ensure that manufacturing data can be compared across production areas, and two to aggregate the manufacturing data to a form stack level.

4.2.3 Data Preparation

The first issue of traceability is addressed by creating a script that utilized the control logic from ISA-88 to follow the individual batch throughout the production processes. The script follows the procedures from the production information management control activity (ISA 1995), where a batch is identified by following the control loops of area, to process control, to unit, and equipment. The script tracks the batch through a process and assign a new ID each time a new production unit is encountered. Because the script is used on historic data, the script is applied backwards, meaning the script starts from the warehouse, where the final product is stored to the milk intake. An example of the logic is depicted in Figure 3. In the example, the traceability script starts from a single pallet and moves backwards in the production, where it will get an id of 'P1'. Then going backwards until it encounters a new production unit, it will receive a new id in the following production unit step, which will respectively be 'P1P1', 'P1P2' for the pieces level, and 'P1P1M1', 'P1P2M1', 'P1P1M2', 'P1P2M2' for the maturity stack level, and continue until there is no more new production units. The process ends when the traceability script is run on all selected batches. The presented example in the figure, is not a full representation of all traceability values for a pallet, as it becomes too big to visualize, where a limitation of two levels per production unit and one level for pieces and maturation stack, is applied.

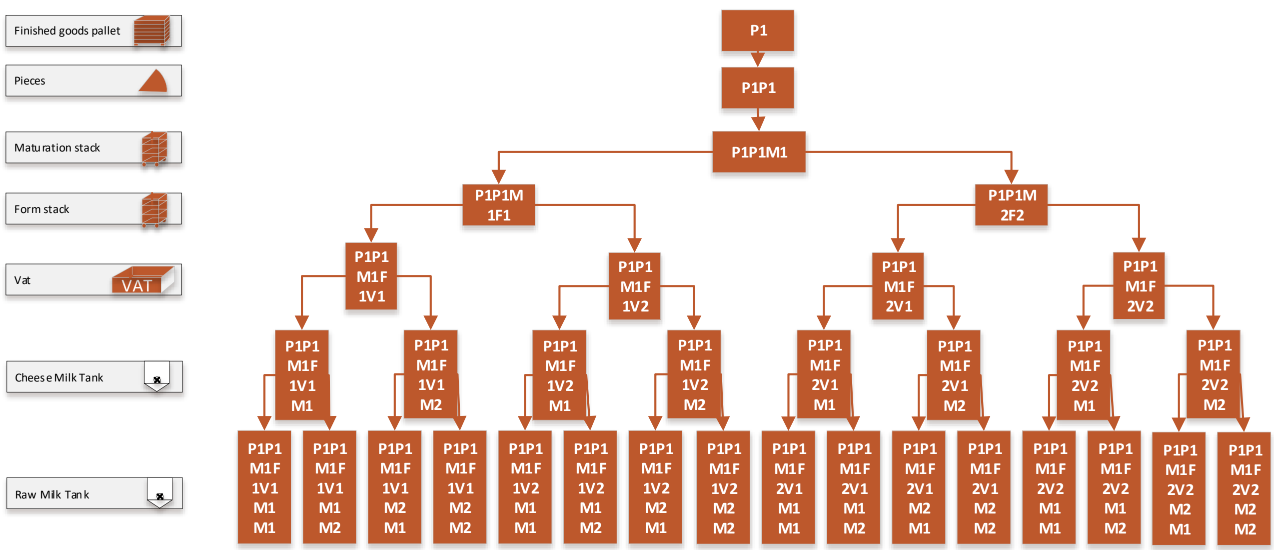
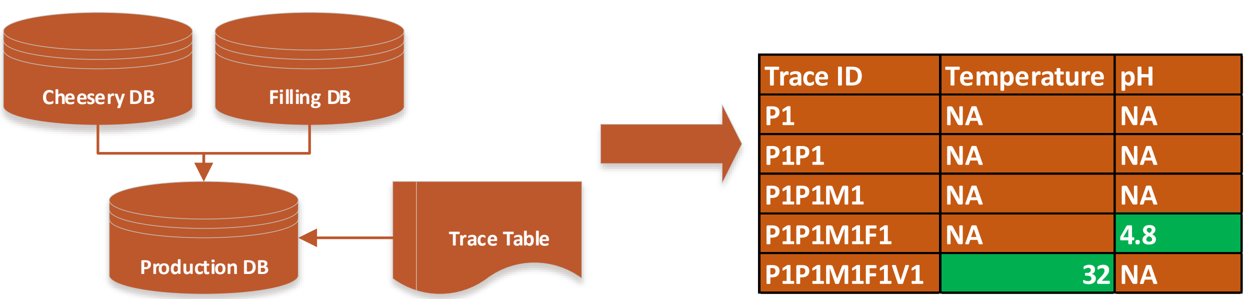


Figure 3 - Example of the traceability script

Executing the script saves the traceability values into a new table in the production database, which is used to uniquely identify all manufacturing data for a pallet, and thus ensures full traceability of the manufacturing data. Next step is to join the databases based on the batch id from ISA-88 and add the corresponding traceability id from the newly created traceability table. Once that task is complete, the data is transformed into a 'tidy' dataset (Wickham 2014), where an example is shown in Figure 4.



Trace ID	Temperature	pH
P1	NA	NA
P1P1	NA	NA
P1P1M1	NA	NA
P1P1M1F1	NA	4.8
P1P1M1F1V1	32	NA

Figure 4 -Example of tidy data and join of data

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

The next challenge for the data preparation stage is to aggregate the data for each production unit level. As can be seen in Figure 4, there are many NA values, where the process values are only stored on the production unit level for which it has been recorded. In the example, the pH value is recorded on the VAT level and temperature on the form stack level. The data is aggregated based on the traceability hierarchy established by the traceability id. The aggregation will take a summary metric, such as the mean, median, maximum, or minimum of the values, when moving up in the hierarchy and the values will be inherited when going down in the hierarchy. An example is presented in Figure 5, where the pH values recorded at the form stack level, is inherited to the VAT level, and the mean of the pH values in the form stack is aggregated to higher levels.

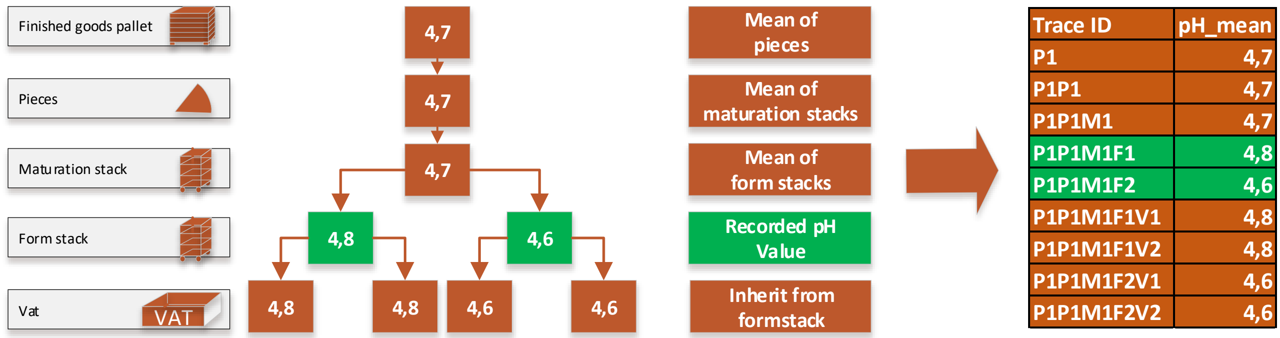


Figure 5 - Example of aggregation by the mean for pH values

The aggregation is performed on all selected process data, for all production unit levels and for the summary metrics mean, median, maximum and minimum.

Having created full traceability on a production unit level, as well as aggregating and 'tidying' the data concludes the data preparation stage.

4.2.4 Modelling

The purpose of the modelling stage is to address the business problems identified in the business understanding phase, by creating an analytic artefact based on the transformed manufacturing data. As there are three business problems to address the analytic model will be based upon three different analytical methods. To summarise the analytical artefact must provide visualisation of the correlations between the production processes, construct a predictive model to predict the 24-hour pH value, and construct a prescriptive model to recommend which actions to take to reach a desired pH level. The selected analytical approach must be using a supervised regression model, as the target variable is known and is non-categorical. Each business problem will now be addressed sequentially in relation to how they can be addressed by an analytical method. Further, it is also important to design an evaluation test for the model to ensure reliability and high performance, which is also known as monitoring the variability and bias of the model.

4.2.4.1 Identifying analytical models

A trained analytical model can often be considered a black box, where the complexity and underlying logic of the model is too difficult for a person to understand. This signifies a need to unlock the black box, where the underlying logic of the model can be presented in a form consumable for a worker at the dairy. The field of xAI are working on ways to simplify and visualise the complexities of a model, enabling the possibility to extract information from a trained model. The methods within xAi is rapidly evolving and are currently able to interpret linear, logistic, decision trees, naïve bayes, and k-nearest neighbours (Molnar 2019), and some experimental research shows that xAi can be used on neural networks as well (Lundberg et al. 2020). However, many of the xAI methods are not consistent in assigning feature importance (Lundberg, Erion, and Lee 2018; Lundberg and Lee 2017), which essentially means the importance of features can change when a model is retrained on new data, which can possible contradict old models, and consequently loose the trust of the user of the model. As an analytical model would likely be retrained during its lifetime, means it is essential, that the presentation of the underlying logic does not change from each time the analytic model is retrained. The SHAP method, however does not suffer from these issues, as it is both globally and locally interpretable, as it is based on additive feature attributes. SHAP is based on shapley values which is the only method which is based on a solid theoretical foundation from game theory (Molnar 2019). The predictions from SHAP is fairly distributed between feature values and enables contrastive explanations, which means the effect of changing a single value can be evaluated. The use of SHAP enables the visualisation of interaction, importance, dependence and clustering of features. As calculating shapley values are computationally heavy, a package is developed that speeds up the calculations, but limits the available analytical models to XGBoost, LightGBM, and CatBoost (Lundberg et al. 2020). For this analytical artefact any of the models could be applied, however the XGboost method was chosen, due to the researcher's familiarity with the method. A familiarity with the method was evaluated to be important, as special care is needed to be taken in regards to handling the missing data due to the sampling of data, and in knowing how the model interacts with correlated features.

Having identifying the requirements for the first two business problems, leaves the last task of identifying an analytical method to create a prescriptive model. Prescriptive models are conceptually optimization or simulation methods, which usually will find an optimal solution based on a fitness function. It can be difficult, though not impossible, to create such a fitness function based on statistics, as it requires a high degree of domain and statistical knowledge to succeed. While this can be done by workers with the right capabilities, the approach is costly to construct and can be difficult to scale. Instead what is needed is a prescriptive model that can be trained and learns on data. In many cases this entails the use of a machine learning algorithm. Many machine learning algorithms can be used to address such an issue. However, it can be difficult, before use, to select the best method. Usually, this should be meet with benchmarking different machine learning algorithms, however the aim of the research is to identify the mechanisms for integrating analytics with MES, where it is found that an optimal selection of machine learning method is not important, as a model with a lower error-rate would likely not present new mechanisms. Consequently, the approach is to select a single suitable machine learning method and apply it. The machine learning model genetic algorithm has been chosen for this project, as it can take a predictive model, in this case the XGBoost model, and add a fitness function to identify cases with the lowest overall error rate.

The final analytic artefact is depicted in Figure 6, which addresses all three business problems. The artefact is based upon the XGBoost method, for which all three business problems can be addressed. The benefit of combining the use of a single analytical model with the three business problems, is that a user will understand why the models make predictions or recommendations, as it can be visualised by the use of xAI and it ensures that the models does not contradict each other. Lastly, as the outputs of the analytical models must be usable and presentable for the end user,

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

entails that a dashboard is built as a part of the analytical artefact to integrate the analytical methods.

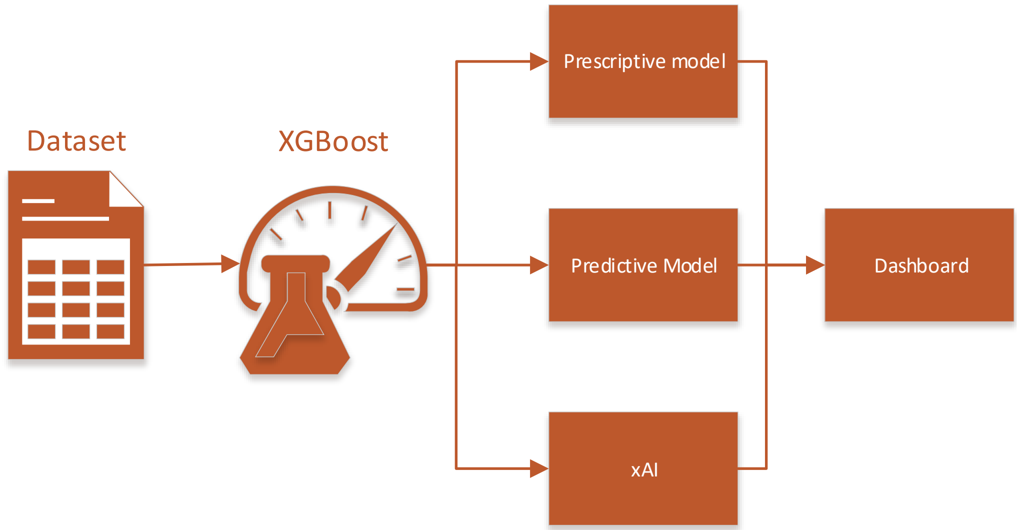


Figure 6 - Analytic artefact

The evaluation of the model will be based on the standard evaluation metric for regression model, which is the root mean square error (RMSE), where the data will be split in a training and test set to avoid the risk of overfitting. As it can be difficult to evaluate what a good error rate is, then a baseline prediction has been made applying a naïve approach, where all predictions are set to the median pH value. This approach produced a RMSE of 0.15

4.2.4.2 Constructing the analytical artefact

The construction of the analytical artefact is built by first training a XGBoost model, then apply the SHAP method on the XGBoost model, and lastly training the genetic algorithm. The overall objective function for all models is to minimize the RMSE in predicting the 24-hour pH value.

The first step in training a model is to select the data to train the model on. In this case there are two parameters that needs to be addressed. One, selecting the aggregation level based on the production units and two, select which summary values to include. As a general rule of thumb, it is nearly always best to choose the aggregation level, where the target variable is stored. In this case, the data is selected on a form stack aggregation level. The data has been aggregated to a mean, median, maximum, and minimum values, where the mean values are removed from the analysis, as there was practically no difference between the mean and median values. The reason for removing the mean values is that having too many covariant features can negatively impact the performance of the model. All available manufacturing data for the cheesery and filling area is selected. The data is then split into a test and training data set, where 80 percent is selected randomly for the training dataset and the remaining data is selected for the test dataset.

Next step in the construction process is to identify hyperparameters for the XGboost model. To identify the hyperparameters a Bayesian optimization method from the R package 'ParBayesianOptimization' is used. The optimization method is applied on the parameters maximum depth, minimum child weight, and subsamples, using a three folds approach on the training data, a learning rate of 0.001, squared error objective function, evaluation metric of RMSE,

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

and is run for 30.000 iterations. The optimal hyperparameters is found to be a maximum depth of 4, a minimum child weight of 18.83369, and a subsample value of 0.34039.

The next step is to run the XGBoost model with the selected hyperparameters. The results of the execution of the model is an RMSE for the training set of 0.057 and 0.069 for the test set. The mean average error is for the training set 0.025, which means that each prediction is on average 0.025 pH value wrong. These results provide a significant better model than the baseline model, where the XGBoost model provides predictions with a 54% lower RMSE. To further evaluate if the model prediction is satisfactory, the prediction model is run on the full data set, where the result is presented in Figure 7.

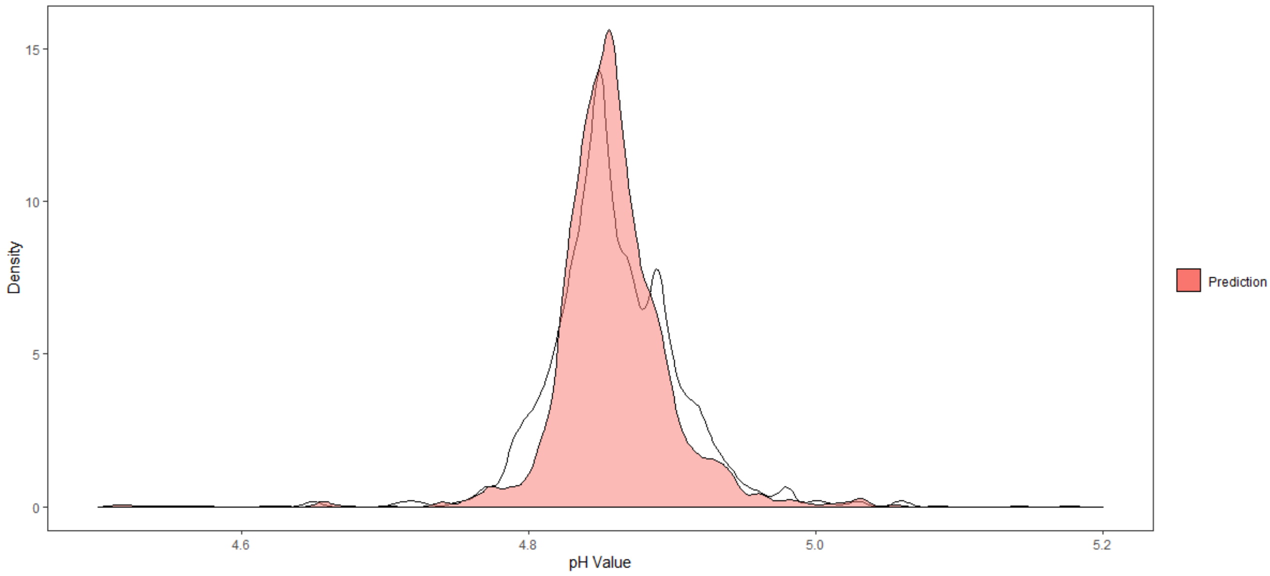


Figure 7 - Predictions of 24-hour pH values

The next task is to interpret the model using the SHAP method. The outcome of the SHAP method is a range of shapley values for each batch, which in itself does not provide a user with much value. The values need to be visualised. The calculations and visualisation of the SHAP values are done by the use of the python package 'shap' and R package 'SHAPforxgboost'. The visualisations provide an understanding of how the different process values for different process steps impact the 24-hour pH value. Two examples are presented in respectively Figure 8 and Figure 9, where the variables names have been anonymised due to confidentiality reasons. The overview plot shows on a summarised level the most important variables that impacts the prediction of the 24-hour pH value in order of importance and plot how the individual values for each variable impact the prediction. As an example, it can be seen that high values for variable 34 will lower the prediction for the 24-hour pH value.

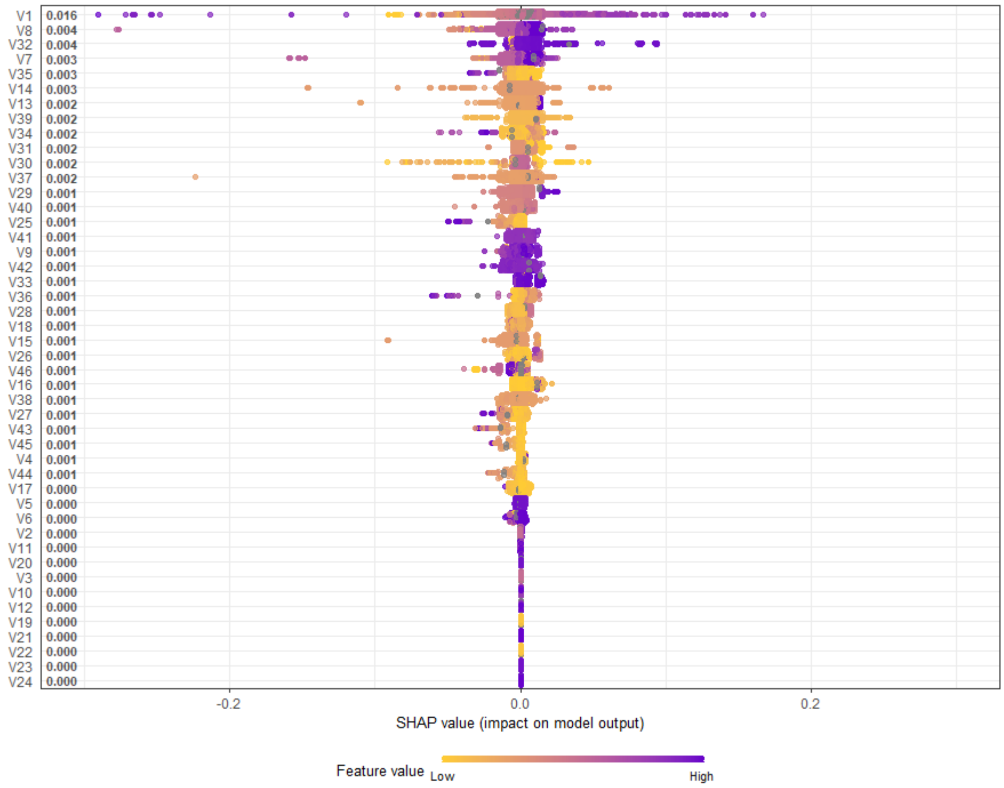


Figure 8 - SHAP Overview

The decision plot shows the impact that each variable for each individual batch have had on the 24-hour pH value. The plots can be used to visualise and create an understanding of the complexity within the manufacturing processes and can identify when and why a 24-hour pH value for a specific batch is high or low.

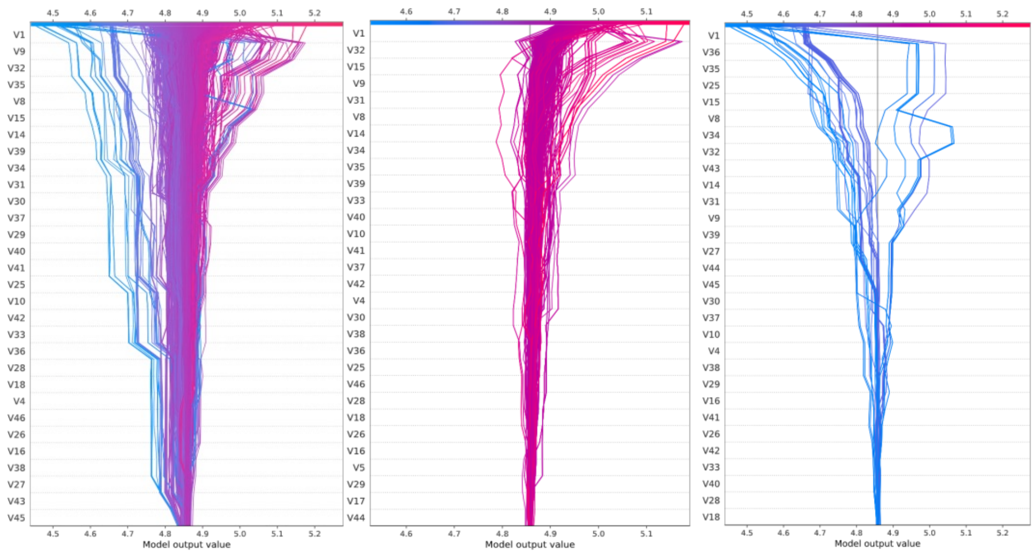


Figure 9 - SHAP decision plots for respectively all (left), high (middle), and low (right) 24-hour pH values for individual batches

Further, a dashboard is built using RShiny, to present all of the plots and enable filtering and selection of data for the individual plot. Additionally, a prediction model from the XGBoost model is also implemented in the RShiny dashboard to enable users to conduct what-if analysis.

The final step of the construction of the analytic artefact is to train the genetic algorithm. To train a genetic algorithm requires two inputs: 1. Fitness function, and 2. selecting hyperparameters. The purpose of the genetic algorithm is to provide recommendations for production values for each production variable in the data. This is realised by reusing the XGBoost model to predict the 24-hour pH value in different scenarios, as a simulation, and compare that with the desired pH value. Thus, the genetic algorithm simulates different predictions made by applying the XGBoost model, and progressively learning which combinations of the variable values results in the desired pH-value. Thus, the fitness function must meet these requirements.

The fitness function for this analytic artefact reuses the trained XGBoost model, by taking the absolute difference between a prediction made by the XGBoost model and the desired pH-value. The genetic algorithm will insert different values for each variable, which is processed by the XGBoost model and calculate an error. The solution space for the genetic algorithm is limited by the minimum and maximum values from the manufacturing data for each variable. Further, as a result of the aggregation of the data, means that each variable has a median, minimum, and maximum value. To ensure that the recommended median and minimum value are not higher than the maximum recommendation, and that the maximum and median recommendation are not lower than the minimum has a penalty been inserted into the fitness function. The fitness function is presented in pseudo code below, where it should be noted that the genetic algorithm always maximises the function, for which the score is transformed to a negative number to practically minimize the error.

$$
\begin{aligned}
&Fitness\ function = (\\
&\quad error = -abs(desired\ pH\ value - xgboost\ prediction)\\
&\quad if\ max < median\ then\ error = error * 2\\
&\quad if\ \min > median\ then\ error = error * 2\\
&return\ error)
\end{aligned}
$$

The following hyperparameters has been selected for the genetic algorithm: population size=92, elitism=0.1, probability of mutation=0.2, probability of cross over = 0.8, and with a mutation based on the normal distribution. The hyperparameters has been selected to ensure that a large part of the solution space is used, as to ideally avoid locally optimised solution, and to simulate different batch production scenarios. The algorithm is run for 50.000 generations, where the results are depicted in Figure 10.

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

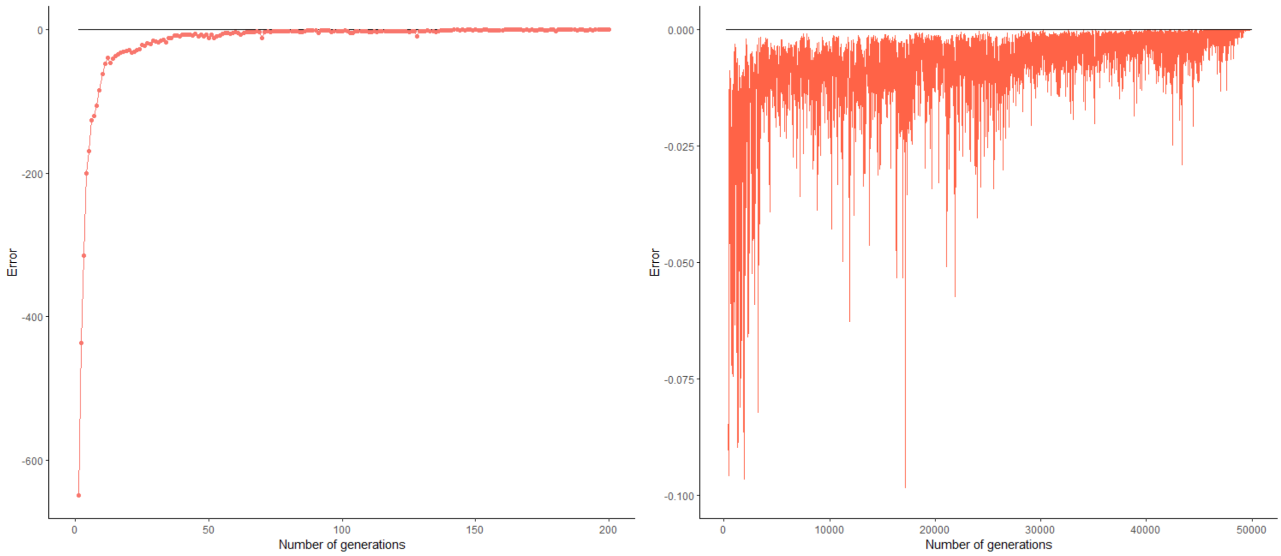


Figure 10 - Genetic algorithm results

The figure visualises a simulation of 50.000 different batch productions and the error for the individual batch in relation to the desired 24-hour pH value. The batch with the lowest error can be considered the optimal way of producing a batch in relation to reaching the desired 24-pH value. However, it can be difficult, in a real-world setting, to produce a batch in exactly the same way for every batch. Therefore, the recommendation model must provide actions for batches, which have not been following an optimal route, to get as close to the desired 24-hour pH value as possible. The simulation is transformed into a recommendation model by identifying the similarity between a batch being produced and the 50.000 simulated batches, calculated as a percentage difference. The recommendation for actions is based on a weighted criterion, where the similarity between the batches and the error is multiplied with each other, and the simulated batch with the lowest weighted error is recommended. By applying this model, the workers at the manufacturing site can get recommendations for actions on a batch level, which updates as the production of a batch progresses. The performance and usefulness of the genetic algorithm can be difficult to evaluate, where in this case the trustworthiness is based upon the use of and inclusion of the XGBoost model. The trained genetic algorithm identified tens of thousand solutions, where the predicted 24-hour pH value is less than 0.025 pH value from the desired pH value. Thus, it is evaluated that the genetic algorithm succeeded in providing recommendations for how to reach a desired pH-value. The genetic algorithm model is implemented in the dashboard alongside the other models, which concludes the construction of the analytic artefact.

4.2.5 Evaluation

The evaluation part of the CRISP-DM framework is not to evaluate the performance of the model, as that was part of the model building phase, but to evaluate if the outcome of the model building phase sufficiently addressed the business problem.

To iterate, the goals was to construct an analytic artefact with a predictive and prescriptive model and present visualisations of the correlation of the production processes impact on the 24-hour pH value. The analytic artefact was successfully constructed, as it addressed all three business problems. Further, the analytic artefact was implemented into a dashboard, which enabled both user and researchers to evaluate the artefact.

4.2.6 Deployment

The deployment of the analytic artefact is made in the cloud on a virtual machine, where manufacturing data from the MES-system can be loaded, transformed, and analysed. The dashboard is also deployed on the virtual machine, for which all relevant user and participants have access to.

4.3 Summarised findings

This section presents the summarised findings and learnings in the construction of the analytic artefact in regards to identifying the mechanism of integrating analytical artefacts and MES.

The main finding is that it was possible to construct an analytical artefact by only addressing the data model. The specific processes, people, IT-systems, and tasks of the manufacturing site were not considered in the construction of the artefact and relied solely on the manufacturing data formatted to the ISA-88 standard. While the users of the manufacturing site identified the business problem, they had no impact on the construction of the artefact. The construction of the artefact was built based on the business problem identified and the transformed manufacturing data. This finding indicates that it is possible to implement the same artefact in another environment with other IT-system, processes, people, and task, as long as the data model is similar.

While the findings indicate that a similar data model is required to implement the analytic artefact in another environment, it was also found that several transformations of the data was required. The data from the manufacturing site follows the ISA-88 standard which is designed for execution and control of production processes (ISA 1995), but not for analytical purposes. To transform the data to also support the use of analytical models, required the data to be traceable on a production unit level, as well as aggregating the data on a production unit level.

The modelling approach was based on the analytic paradigm of extracting relationships and correlation from the data instead of creating purpose built statistical models. In practice the application of the analytical artefact showed that off-the shelf methods can be used on the manufacturing data to get a good analytical model. While there are many other factors for obtaining and securing value through analytical models, such as change management, these findings indicate that from an IT perspective, it is possible to scale an analytic artefact across heterogeneous manufacturing sites to solve similar business problems.

Further, extrapolating the findings indicate that an analytical artefact can be built, which can be deployed to several heterogeneous manufacturing sites, if the data is traceable and aggregated on a production unit level. This concept is visualised in Figure 11, which essentially shows the mechanism needed to integrate analytics with MES, which will be evaluated in chapter 5.

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

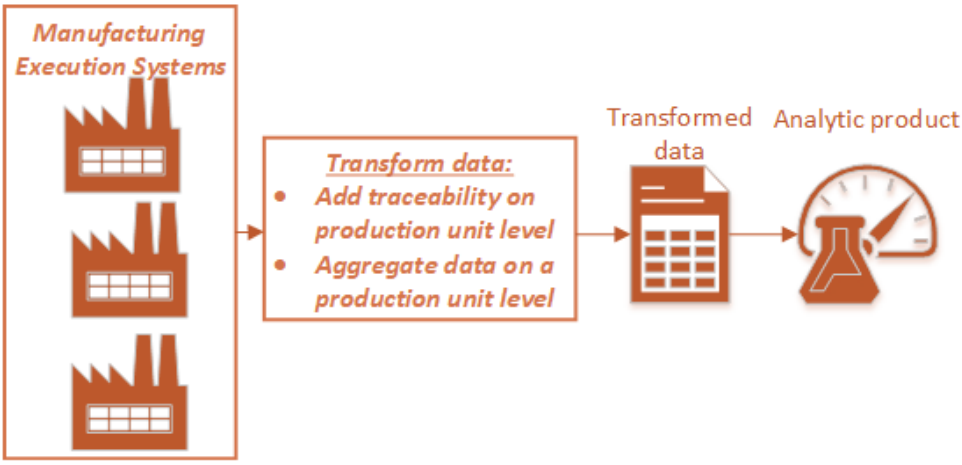


Figure 11 - Analytic product concept

A part of applying the design science methodologies, is that the design process can present unintended findings. An unintended finding in the development and construction of the analytic artefact, are two types of barriers for maturing the use of analytics. The barriers for maturing the use of analytics are respectively, a data barrier which must be passed to enable the use of diagnostic analytics, and an analytical expertise barrier which limits the ability to apply predictive and prescriptive analytics. The barriers are visualised in Figure 12, where the data barrier limits the use of diagnostic, predictive, or prescriptive analytics, and the analytic expertise barrier should be passed before predictive or prescriptive analytics can be applied. The data barrier in this case was passed by the transformation of data to include traceability and aggregation of manufacturing data on a production unit level. The analytical expertise barrier is understood as the need for human knowledge and capabilities within applying predictive and prescriptive analytical models.

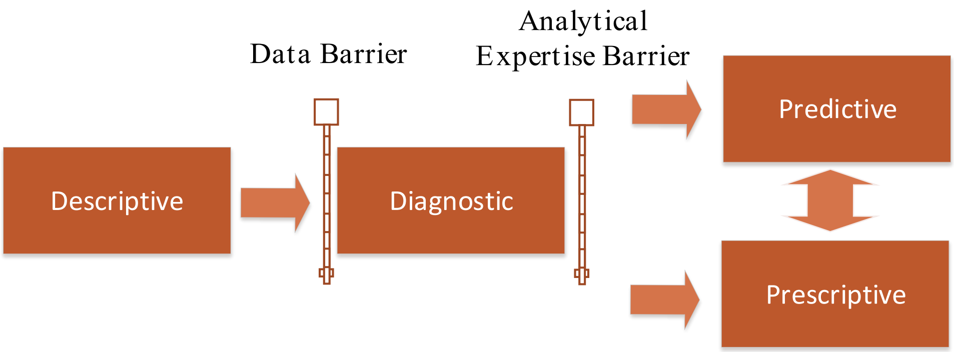


Figure 12 – Barriers for maturing analytics

As a consequence, it is possible to identify the capabilities needed for maturing the use of analytics, which is presented in Table 1. The table should be read from left to right, where e.g. the reason for a low score for the data management capability for the predictive maturity stage is that the data has already been transformed in the diagnostic stage. These findings show that as companies are implementing BI solutions, they are unintendingly preparing for the use of more advanced levels of analytics. For these companies to mature to the predictive or predictive stage requires people with analytical capabilities to develop analytical models alongside domain experts.

	Descriptive	**Diagnostic**	**Predictive**	**Prescriptive**

Data management capability	Low	High	Low	Low
Analytical capability	Low	Low-Medium	Medium	High
Domain expertise	Low	Low	Medium	Medium

Table 1 - Capabilities needed for maturing the use of analytics

The software used are all open source, which are the statistical programming language R, python and Microsoft SQL server management. The cost for the project consisted of the working time for the researchers and for the cost of hosting a virtual machine. However, the project could have been conducted on a standard issue laptop, removing the cost of the virtual machine. The project shows that by leveraging the analytic capabilities of open source software, enables the ability to load and process the data from the MES system into advanced analytical models.

5 Evaluation of findings

This section will evaluate the findings from the previous section by deploying the artefact into another environment, i.e. another manufacturing site. The deployment will transform the data model of the manufacturing site into a shared data format, and as a result will deploy the artefact without considering other parameters such as different IT-systems, process, people, and tasks. As the artefact is already constructed there is no need to use the CRISP-DM framework.

The second manufacturing site is a large dairy within the same company, who produces mozzarella cheeses, which is also classified as a batch type production. The manufacturing processes of the production site is mapped in Figure 13, according to the same standards as figure 1.

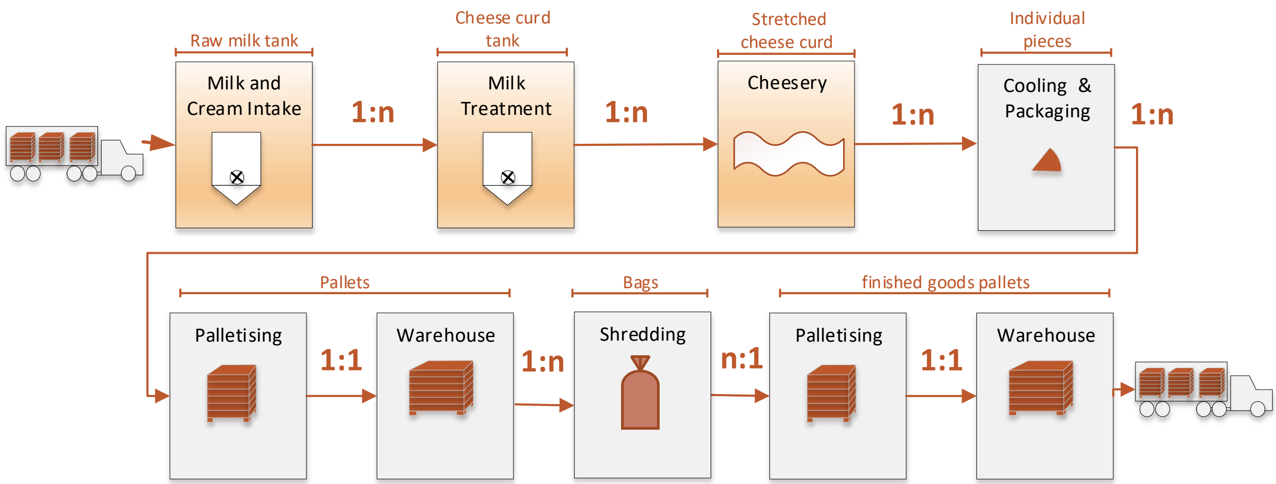


Figure 13 - Overview of the second manufacturing site

The manufacturing of mozzarella cheeses is very different than making blue cheeses, where producing mozzarella is significantly faster and goes through fewer manufacturing steps. The manufacturing site have a different IT-system landscape compared to the first manufacturing site, and while they also apply the ISA-88 standard, they do so in a different way as e.g. the naming conventions is different. All data is stored in a single database, meaning control, execution and manufacturing data is stored alongside each other, with a batch ID to join the process data. However, there is no traceability id on a production unit level. The process data is collected by

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

sampling however, the manufacturing site have a higher sampling rate, compared to the first manufacturing site.

5.1 Preparing data for the analytic artefact

As the analytics artefact is already constructed, the only pre-processing step needed is to create traceability on a per production unit level and aggregating the data. A traceability script is created using the same logic as for the first manufacturing site, where the script had to be rewritten from scratch, as the implementation of the ISA-88 standard was different. Writing the script took approximately two days. Next, all of the manufacturing data was aggregated to each production unit level, using the same logic as for the first manufacturing site, which took about an hour to complete. Having finished the two data preparations steps, concludes the data preparation step.

5.2 Applying the analytic artefact

To apply the analytic artefact means that the XGboost model must be retrained on new data, as well as identifying hyperparameters. Further, the shap method should be executed based on the new data and trained XGboost model, as well as executing the genetic algorithm on the new data. The XGboost model will be trained to predict the pH values recorded in the cheesery production area and include all data from the previous production areas.

The first step of applying the analytic artefact is to identify hyperparamters for the XGboost model. The hyperparameters are found by using the same function from the R library '*ParBayesianOptimization',* where the optimal hyperparameter values have been found to be: learning rate of 0.001, maximum depth of 7, minimum child weight of 18.83369, a sub sample rate of 0.34039 and evaluated by the RMSE. The manufacturing data is split into a training and test set, by respectively randomly selecting 80 and 20 percent of the manufacturing data.

The XGboost model is trained on the training data, with an RMSE on the test data of 0.16 and 0.11 for the training data, and a score of 0.12 for the mean average error. This needs to be compared with the naïve prediction using the median prediction method, which resulted in an RMSE of 0.31, which is an improvement of 49%. Finally, to visualise the results of the prediction model, the model is used on all manufacturing data, where the result is presented in Figure 14.

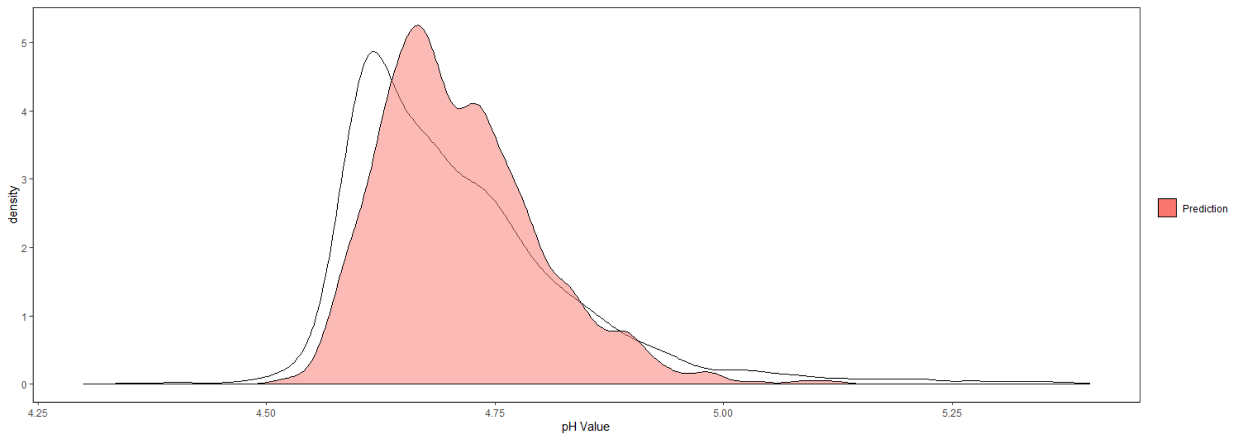


Figure 14 - Predictions for the pH value for the second manufacturing site

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

The next step is to calculate the shap values and visualise the results. The shap values are calculated by the use of the python package 'shap', where an overview plot is presented in Figure 15, and a decision plot is presented in Figure 16.

The manufacturing site can use the overview plot to e.g. identify that high values of variable 49 will lower the predicted pH value, and high values of variable 58 will make the prediction of the pH value higher.

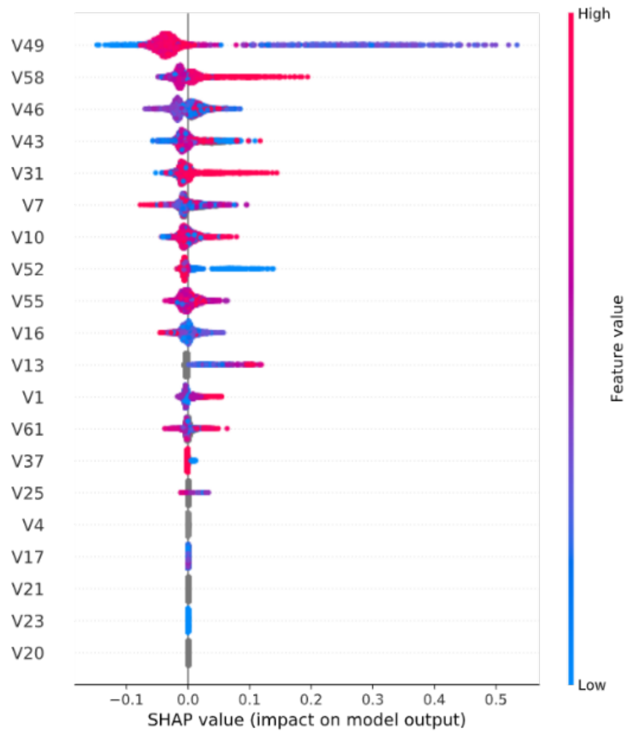


Figure 15 - SHAP overview plot for the second manufacturing site

The decision plots show that there are distinct patterns for the individual production batch, for reaching a high or low pH value. Specifically, it seems that once the variables indicate a low pH value, the end result is usually a low final pH value. Knowing this, the workers of the manufacturing site can make sure to change the manufacturing of the cheese batch to adjust for a higher pH value.

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

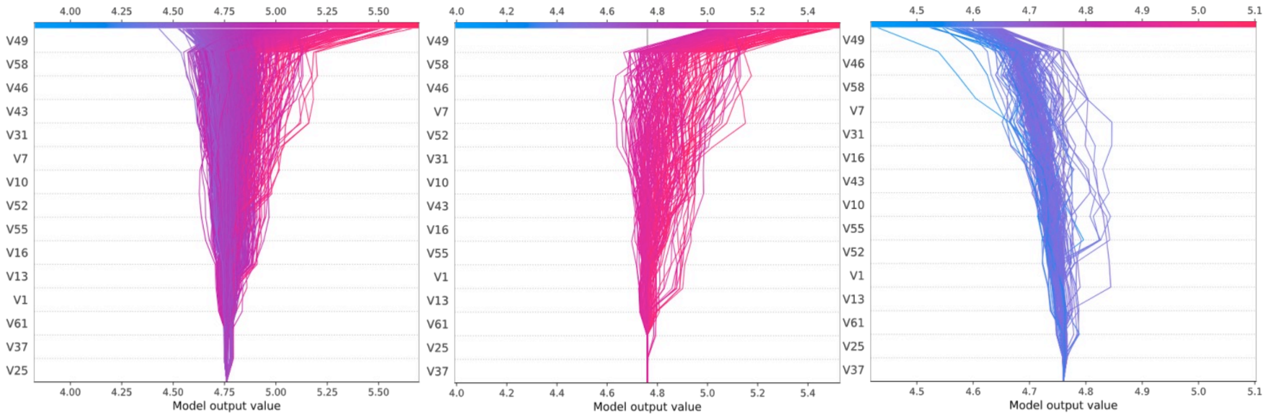


Figure 16 - SHAP decision plots for the second manufacturing site, respectively all (left), high (middle), and low (right) pH values for individual batches

The final step in applying the analytical artefact is to apply the genetic algorithm. The application of the algorithm is fairly simple, as both the fitness function and hyperparameters have been created and selected. Essentially, this leaves the simple task of execution the genetic algorithm on the new manufacturing data and the newly trained XGboost model, which is done 50.000 times and the results are depicted in Figure 17.

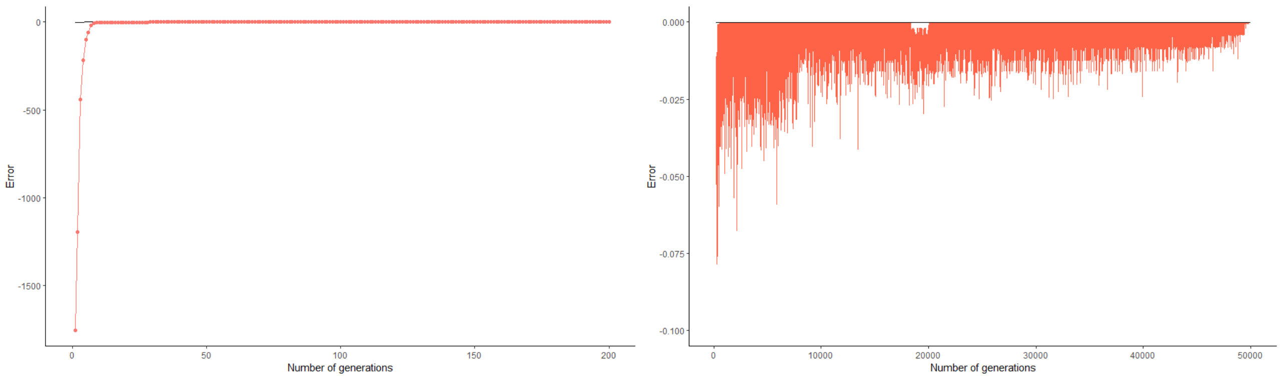


Figure 17 - Genetic algorithm results for the second manufacturing site

Lastly, all of the methods are implemented into a dashboard, which is deployed onto a virtual machine where both users and researchers have access to.

5.3 Summarised findings

The purpose of this section is to evaluate if the findings and learnings from the construction of the analytics artefact, can be reproduced in another environment. The findings that are addressed in the evaluation are threefold. One, can an analytical artefact be deployed by only transforming the manufacturing data to a common data model, two, is there a data and analytical expertise barrier, and three, are the same capabilities for data management, context knowledge, and analytical capability needed for the deployment of an analytic artefact.

The first finding is considered verified, as the analytic artefact is applied successfully by only transforming the manufacturing data into a common data model, without considering other parameters such as it-systems, people, processes, or tasks. The total time of applying the analytic artefact to the second manufacturing site took about three days in total, by a single person, which indicates a great potential for scalability. The finding also confirms the possibility to construct an analytical model to address a specific business problem, centrally and deploy the artefact to multiple heterogeneous manufacturing sites.

The application of the analytical artefact showed that there are different requirements for capabilities between constructing and deploying an analytical artefact. While the same degrees of data management capabilities were needed for deploying the artefact, due to having to create a new script to ensure traceability and aggregate the data on a production unit level, the same levels of domain and analytical capabilities where not needed. In fact, all of the analytical applications can be automised, by e.g. auto-tuning hyperparameters, automate the calculation and visualisation of shap values, automatically executing the genetic algorithm, and deploying the results to a dashboard. While having domain knowledge is not technically needed for applying the analytical artefact, it is likely to be essential for extracting value out of the analytic artefact, to ensure that the outcome of the artefact is business relevant. As a consequence, the data barrier also exists for deploying an analytical artefact, however there is no to little need for analytical expertise in deploying an analytic artefact.

6 Discussion

The aim of the paper was to investigate the mechanism for integrating MES with an analytical artefact, which was achieved by designing, constructing and deploying an analytical artefact in two manufacturing sites. The outcome is design knowledge which have led to both managerial and research findings, which is discussed in this section.

6.2 Managerial Implications

6.2.1 Integrating an analytic artefact

Integrating an analytical artefact is very different than integrating EIS solutions into a company. Unlike EIS integrations, an analytical artefact can be constructed based on the available enterprise data, human capabilities, and open source software. The integration of an artefact does not need to be a big and costly project, but can be done by few people and be deployed as a microservice. As such there is no need to replace the software a company uses, as an analytical artefact can simply read the available data through an API and thereby provide insights and actions. In fact, the application of an analytical artefact, is not limited to a manufacturing context, but can be constructed to be applicable for most planning and execution processes for the supply chain company. An analytical artefact could provide better and more detailed insights and actions in e.g. production scheduling or demand planning than what is offered in the current MES, ERP or APS offerings. The current EIS offerings are often tailored toward reporting or executing purposes and in most cases lacks access to predictive or prescriptive analytical methods. On the other hand, open source software such as R and Python, both have access to state-of-the-art analytical methods, which are updated at least weekly. While EIS do not possess much capability within advanced analytics, they are very capable of ensuring good planning and executing processes. As a consequence, there is a great potential in integrating analytical artefact with EIS, by improving the planning and execution activities and provide additional insight into the manufacturing and supply chain complexities by the use of analytics. The integration can be achieved in many ways, but as the with the rise of cloud technology, a likely deployment strategy would be by the use of microservices often deployed in a containerised environment.

This is an Accepted Manuscript of an article published by Taylor & Francis in Enterprise Information Systems on 8 April 2021, available online: https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1905881

To enable a deployment of an analytical artefact for several manufacturing sites, it was found that two prerequisites must be fulfilled, namely having traceability and aggregation on a production unit level. The case company in this instance used ISA-88, and the scripts for constructing the artefact was based on that standard. However, the logic used for creating both the traceability and aggregation scripts can likely be used in manufacturing sites with other data standards. The script constructed for the analytical artefact would likely also be applicable for ISA-95, as ISA-88 and ISA-95 to a large degree share terminology and aims (Scholten 2007). However, the scripts for creating traceability and aggregation, should be improved in regards to speed and resource usage, if it is to be integrated into a production environment.

While it is recommended to deploy the analytical artefacts as microservices for companies, they do present their own set of challenges. With an increase of microservices requires a company to manage these microservices, which can potentially present a cost greater than the combined cost of constructing and deploying an analytical artefact (Sculley et al. 2015). While the construction and deployment of the analytical artefact was conducted within batch production manufacturing sites, it can be assumed that the findings are also relevant for other types of manufacturing.

6.2.2 Capabilities

It was found that human capabilities are essential for constructing an analytical artefact, where it is found that analytical, data management, and IT capabilities as well as domain knowledge is required. Further, constructing and deploying analytical artefacts require different capabilities, as well as different capabilities were needed for the different construction phases.

The first phase of constructing an analytical artefact, i.e. understanding the business and data, mostly relies on domain knowledge and analytic capabilities to identify a relevant business issue and identify analytical models to solve the business problem. The data preparation and modelling stages requires data management, IT, and analytic capabilities. The deployment phase did not require the use of domain knowledge or analytical capabilities, and can mainly be seen as an IT task. As a consequence, it is likely that the deployment of an analytical artefact can be automatically deployed to many sites without human involvement, which presents an opportunity to scale an artefact at a low cost. Managing the construction and deployment of an analytical artefact, will be different, as e.g. the construction can be made centrally by a team consisting of analytical expert and data scientist, which then is deployed by the local IT departments of the heterogeneous manufacturing sites.

In addition, it is found that analytical artefact can be constructed to answer generic business problems. In this case, an artefact is constructed to explain, predictive, and recommend actions based on the correlations between manufacturing processes. The artefact constructed in this paper uses the target variable 24-hour pH value, however almost any variable within the manufacturing site could have been chosen as the target variable. Consequently, introducing a graphical user interface and auto-ML functionalities would enable users with limited IT and analytical capabilities to use an artefact to solve their specific business problem. This could lead to a user focused self-service analytical solution, where the application of advanced analytics is democratised to the users.

6.3 Research implications

The research domains of EIS and SCM have had a lot of research interest in regards to understanding the mechanism for a successful implementation and assimilation primarily by researching critical success and failure factors (Nazemi, Tarokh, and Djavanshir 2012; Stefanou 2001; Koh, Gunasekaran, and Goodman 2011). However, there has been little research interest into how analytics can be integrated with EIS. This paper provides insight into the mechanisms needed to integrate MES and analytics from an IT perspective, but does not address many other aspect of integrating analytics into a company. It is uncertain which critical factors there is for implementing and integrating MES and analytics, but especially top-management support, user training, performance evaluation, and context depend configuration is highlighted as potential critical factors.

This paper finds that the integration of an analytical artefact and MES is different than a typical EIS implementation. The construction, deployment and integration processes are much less intrusive for the company, and can be constructed and deployed without interrupting daily operations and can be conducted at a fairly low budget. The processes of integrating analytics with EIS, is in its entirety different where an EIS implementation usually consist of purchasing a standard software product from a vendor, and customizing the product to fit to the company. On the other hand, integrating analytics consist of constructing a small piece of software, using open source software tools, to address a specific business need, which can be accomplished at low cost, in a short time with a few people. The scope of the implementations is therefore vastly different and as a consequence, must be treated differently, which presents an opportunity for researchers to address this particular issue.

The mechanism for integrating analytics and EIS has for this paper been researched within the scope of MES, however it is believed that the findings are generally applicable for all EIS. Though, there is still a need to identify how specifically analytics can improve the planning and execution processes of EIS. While the research fields of SCM, analytics, and EIS is often regarded as an applied science, most of the research produced are not relevant for practice (Jonsson and Holmström 2016). As a consequence, there is a need to identify the mechanisms of how the many analytical and OR models produced by researchers can be integrated into EIS environments. If this is accomplished and the mechanism are understood, could potentially lead to an analytical revolution within the domain of SCM. The research in this paper aids in that regard in closing the gap, as it provides empirical research to a field which primarily has conceptual research (Asmussen and Møller 2020).

7 Conclusion

This paper investigated the mechanisms for integrating an analytical artefact with MES by the use of the ADR methodology. The analytical artefact was constructed and deployed to a single manufacturing site, and later deployed to a second manufacturing site. The construction and deployment of the analytical artefact have led to the prescriptive generation of design knowledge, which have identified mechanisms for integrating MES with an analytical artefact from an IT-perspective.

The findings in this paper show that the current implementation of MES, which are implemented for control and execution purposes, can be transformed to also incorporate advanced analytics, by transforming the data model and process the transformed data by the use of state-of-the-art analytical models found in open source software.

It is found that analytic artefacts can be constructed and deployed based on a business problem, where the specific context of the manufacturing site does not matter, i.e. an artefact can be constructed without considering the IT-systems, processes, task, and products of the manufacturing site. However, to integrate the artefact with MES requires the manufacturing data to be transformed into a data model which has traceability on a production unit level and is aggregated on a production unit level. If these requirements for the data model is fulfilled, then the artefact can be deployed at scale across heterogeneous manufacturing sites.

Constructing the analytical artefact is restricted by two barriers. One, is a data barrier, which must be passed to enable diagnostic, predictive and prescriptive analytics, and an analytical capability barrier which is a requirement for enabling predictive and prescriptive analytics. Constructing an analytical artefact using advanced analytics require that the data is available in the right formats, and that there are human capabilities to construct an artefact based on open source software. It is found that the construction and deployment of an analytical artefact is different than IT or EIS integrations, where the integration of an analytical artefact is smaller in scope, faster to integrate, and require a different set of skills to construct, i.e. analytical, IT, and data management skills, as well as domain knowledge. Further, it is found that the deployment of an artefact can be treated as an IT integration project, where the artefact can be integrated into a new manufacturing site, without any analytical skills by automating the execution of the artefact and the use of auto-ML. Further, introducing a graphical interface to the artefact would enable non-analytical experts to use the artefact for advanced analytics, effectively creating a scaling self-service advanced analytics artefact.

Declaration of interest statement

None

Bibliography

Addo-Tenkorang, Richard, Petri T. Helo, and Jussi Kantola. 2017. “Concurrent Enterprise: A Conceptual Framework for Enterprise Supply-Chain Network Activities.” *Enterprise Information Systems* 11 (4). Taylor & Francis: 474–511. doi:10.1080/17517575.2016.1212274.

Akyuz, Goknur Arzu, and Mohammad Rehan. 2009. “Requirements for Forming an E-Supply Chain.” *International Journal of Production Research* 47 (12): 3265–87. doi:10.1080/00207540701802460.

Appelbaum, Deniz, Alexander Kogan, Miklos Vasarhelyi, and Zhaokai Yan. 2017. “Impact of Business Analytics and Enterprise Systems on Managerial Accounting.” *International Journal of Accounting Information Systems* 25 (March). Elsevier: 29–44. doi:10.1016/j.accinf.2017.03.003.

Asmussen, Claus Boye, and Charles Møller. 2020. “Enabling Supply Chain Analytics for Enterprise Information Systems: A Topic Modelling Literature Review.” *Enterprise Information Systems* 14 (5). Taylor & Francis: 563–601. doi:10.1080/17517575.2020.1734240.

Barbosa, Marcelo Werneck, Alberto de la Calle Vicente, Marcelo Bronzo Ladeira, Marcos Paulo Valadares de Oliveira, and Marcos Paulo Valadares de Oliveira. 2017. “Managing Supply Chain Resources with Big Data Analytics: A Systematic Review.” *International Journal of Logistics Research and Applications* 21 (3). Taylor & Francis: 177–200. doi:10.1080/13675567.2017.1369501.

Beer, David. 2018. “Envisioning the Power of Data Analytics.” *Information Communication and Society* 21 (3). Taylor & Francis: 465–79. doi:10.1080/1369118X.2017.1289232.

Benaissa, Mounir, and Abdellatif Benabdelhafid. 2005. “SOAP Based System for Integration in the Logistic Chain.” *Journal of Decision Systems* 14 (1–2): 55–82. doi:10.3166/jds.14.55-82.

Bose, Ranjit. 2009. “Advanced Analytics: Opportunities and Challenges.” *Industrial Management & Data Systems* 109 (2): 155–72. doi:10.1108/02635570910930073.

Breiman, Leo. 2001. “Statistical Modeling: The Two Cultures.” *Statistical Science* 16 (3): 199–215. doi:10.1214/ss/1009213726.

Carillo, Kevin Daniel André. 2017. “Let’s Stop Trying to Be ‘Sexy’ – Preparing Managers for the (Big) Data-Driven Business Era.” *Business Process Management Journal* 23 (3): 598–622. doi:10.1108/BPMJ-09-2016-0188.

Cottyn, J., H. Van Landeghem, K. Stockman, and S. Derammelaere. 2011. “A Method to Align a Manufacturing Execution System with Lean Objectives.” *International Journal of Production Research* 49 (14): 4397–4413. doi:10.1080/00207543.2010.548409.

Cupek, Rafal, Adam Ziebinski, Marek Drewniak, and Marcin Fojcik. 2018. “Knowledge Integration via the Fusion of the Data Models Used in Automotive Production Systems.” *Enterprise Information Systems* 13 (7–8). Taylor & Francis: 1094–1119. doi:10.1080/17517575.2018.1489563.

Dai, Hong Ning, Hao Wang, Guangquan Xu, Jiafu Wan, and Muhammad Imran. 2020. “Big Data Analytics for Manufacturing Internet of Things: Opportunities, Challenges and Enabling Technologies.” *Enterprise Information Systems* 14 (9–10). Taylor & Francis: 1279–1303. doi:10.1080/17517575.2019.1633689.

Davenport, Thomas, and Jeanne G. Harris. 2007. *Competing on Analytics: The New Science of Winning*. Boston, Mass: Harvard Business School Press.

Delen, Dursun, and Hamed M. Zolbanin. 2018. “The Analytics Paradigm in Business Research.” *Journal of Business Research* 90 (May). Elsevier: 186–95. doi:10.1016/j.jbusres.2018.05.013.

Demirkan, Haluk, and Dursun Delen. 2013. “Leveraging the Capabilities of Service-Oriented Decision Support Systems: Putting Analytics and Big Data in Cloud.” *Decision Support Systems* 55: 412–21. doi:10.1016/j.dss.2012.05.048.

Gandomi, Amir, and Murtaza Haider. 2015. “Beyond the Hype: Big Data Concepts, Methods, and Analytics.” *International Journal of Information Management* 35 (2). Pergamon: 137–44. doi:10.1016/J.IJINFOMGT.2014.10.007.

Gupta, Shalini, and Erick Jones. 2014. “Optimizing Supply Chain Distribution Using Cloud Based Autonomous Information.” *International Journal of Supply Chain …* 3 (4): 79–90. http://www.ojs.excelingtech.co.uk/index.php/IJSCM/article/view/982.

Hahn, G.J., and J. Packowski. 2015. “A Perspective on Applications of In-Memory Analytics in Supply Chain Management.” *Decision Support Systems* 76 (August). North-Holland: 45–52. doi:10.1016/J.DSS.2015.01.003.

Herden, Tino T. 2017. “Similarity-Matching in Decision-Making Processes of Supply Chain Analytics: A Systematic Literature Review.” *Supply Chain Forum* 18 (1). Taylor & Francis: 13–20. doi:10.1080/16258312.2017.1278670.

Herden, Tino T. 2019. “Explaining the Competitive Advantage Generated from Analytics with the Knowledge-Based View : The Example of Logistics and Supply Chain.” *Business Research*. Springer International Publishing. doi:10.1007/s40685-019-00104-x.

Hindle, Giles A., and Richard Vidgen. 2018. “Developing a Business Analytics Methodology: A Case Study in the Foodbank Sector.” *European Journal of Operational Research* 268 (3). Elsevier B.V.: 836–51. doi:10.1016/j.ejor.2017.06.031.

ISA. 1995. *ANSI/ISA-88.01-1995, Batch Control, Part 1: Models and Terminology*.

Ishaya, Tanko, and Musiliudeen Folarin. 2012. “A Service Oriented Approach to Business Intelligence in Telecoms Industry.” *Telematics and Informatics* 29 (3). Elsevier Ltd: 273–85. doi:10.1016/j.tele.2012.01.004.

Janssen, Marijn, Haiko van der Voort, and Agung Wahyudi. 2017. “Factors Influencing Big Data Decision-Making Quality.” *Journal of Business Research* 70. Elsevier Inc.: 338–45. doi:10.1016/j.jbusres.2016.08.007.

Jeon, Byeong Woo, Jumyung Um, Soo Cheol Yoon, and Suh Suk-Hwan. 2017. “An Architecture Design for Smart Manufacturing Execution System.” *Computer-Aided Design and Applications* 14 (4): 472–85. doi:10.1080/16864360.2016.1257189.

Jiang, P. Y., G. H. Zhou, G. Zhao, Y. F. Zhang, and H. B. Sun. 2007. “E2-MES: An e-Service-Driven Networked Manufacturing Platform for Extended Enterprises.” *International Journal of Computer Integrated Manufacturing* 20 (2–3): 127–42. doi:10.1080/09511920601020664.

Jonsson, Patrik, and Jan Holmström. 2016. “Future of Supply Chain Planning: Closing the Gaps between Practice and Promise.” *International Journal of Physical Distribution & Logistics Management* 46 (1): 62–81.

Kache, Florian, and Stefan Seuring. 2017. “Challenges and Opportunities of Digital Information at the Intersection of Big Data Analytics and Supply Chain Management.” *International Journal of Operations and Production Management* 37 (1): 10–36. doi:10.1108/IJOPM-02-2015-0078.

Khanra, Sayantan, Amandeep Dhir, and Matti Mäntymäki. 2020. “Big Data Analytics and Enterprises: A Bibliometric Synthesis of the Literature.” *Enterprise Information Systems* 14 (6). Taylor & Francis: 737–68. doi:10.1080/17517575.2020.1734241.

Koh, S. C.L., A. Gunasekaran, and T. Goodman. 2011. "Drivers, Barriers and Critical Success Factors for ERPII Implementation in Supply Chains: A Critical Analysis." *Journal of Strategic Information Systems* 20 (4). Elsevier B.V.: 385–402. doi:10.1016/j.jsis.2011.07.001.

Koupaei, Mehrdad Nouri, Mohammad Mohammadi, and Bahman Naderi. 2016. "An Integrated Enterprise Resources Planning (ERP) Framework for Flexible Manufacturing Systems Using Business Intelligence (BI) Tools." *International Journal of Supply and Operations Management* 33 (11): 1112–25.

Kridel, Don, and Daniel Dolk. 2013. "Automated Self-Service Modeling: Predictive Analytics as a Service." *Information Systems and E-Business Management* 11 (1): 119–40. doi:10.1007/s10257-011-0185-1.

Kumar Sangaiah, Arun, Ankit Chaudhary, Chun Wei Tsai, Jin Wang, and Francesco Mercaldo. 2020. "Cognitive Computing for Big Data Systems over Internet of Things for Enterprise Information Systems." *Enterprise Information Systems* 14 (9–10). Taylor & Francis: 1233–37. doi:10.1080/17517575.2020.1814422.

Larson, Deanne, and Victor Chang. 2016. "A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science." *International Journal of Information Management* 36 (5). Elsevier Ltd: 700–710. doi:10.1016/j.ijinfomgt.2016.04.013.

Leng, Jiewu, Jiang, Pingyu, Liu, Chao, and Wang, Chuang, 2020. " Contextual self-organizing of manufacturing process for mass individualization: a cyber-physical-social system approach" Enterprise Information Systems 14(8):1124-1149 https://doi.org/10.1080/17517575.2018.1470259.

Leventhal, Barry. 2015. "All Models Are Wrong but Some Are Useful: The Use of Predictive Analytics in Direct Marketing." *Quality Technology and Quantitative Management* 12 (1): 93–104. doi:10.1080/16843703.2015.11673369.

Link, Björn, and Andrea Back. 2015. "Classifying Systemic Differences between Software as a Service- and On-Premise- Enterprise Resource Planning." *Journal of Enterprise Information Management* 28 (6): 808–37. doi:http://dx.doi.org/10.1108/MRR-09-2015-0216.

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2 (1): 56–67. doi:10.1038/s42256-019-0138-9.

Lundberg, Scott M., Gabriel G. Erion, and Su-In Lee. 2018. "Consistent Individualized Feature Attribution for Tree Ensembles," no. 2. http://arxiv.org/abs/1802.03888.

Lundberg, Scott M., and Su In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 2017-Decem (Section 2): 4766–75.

Mansour, Ikbal, Pierre Alain Millet, and Valérie Botta-Genoulaz. 2018. "A Standard-Based Business Repository for on-Demand Manufacturing Execution Systems." *Production Planning and Control* 29 (7). Taylor & Francis: 602–13. doi:10.1080/09537287.2018.1456691.

Marchand, Donald A, and Joe Peppard. 2013. "Why IT Fumbles Analytics." *Harvard Business Review*, no. 91: 104–12.

Molnar, Christoph. 2019. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*.

Mortenson, Michael J., Neil F. Doherty, and Stewart Robinson. 2014. "Operational Research from Taylorism to Terabytes: A Research Agenda for the Analytics Age." *European Journal of Operational Research* 241 (3). Elsevier Ltd.: 583–95. doi:10.1016/j.ejor.2014.08.029.

Nagorny, Kevin, Pedro Lima-Monteiro, Jose Barata, and Armando Walter Colombo. 2017. "Big Data Analysis in Smart Manufacturing: A Review." *International Journal of Communications, Network and System Sciences* 10 (03): 31–58. doi:10.4236/ijcns.2017.103003.

Nazemi, Eslam, Mohammad Jafar Tarokh, and G. Reza Djavanshir. 2012. "ERP: A Literature Survey." *International Journal of Advanced Manufacturing Technology* 61 (9–12): 999–1018. doi:10.1007/s00170-011-3756-x.

Nguyen, Truong, Li Zhou, Virginia Spiegler, Petros Ieromonachou, and Yong Lin. 2017. "Big Data Analytics in Supply Chain Management: A State-of-the-Art Literature Review." *Computers and Operations Research* 0. Elsevier Ltd: 1–11. doi:10.1016/j.cor.2017.07.004.

Omar, Yamila M., Meysam Minoufekr, and Peter Plapper. 2019. "Business Analytics in Manufacturing: Current Trends, Challenges and Pathway to Market Leadership." *Operations Research Perspectives* 6 (October). Elsevier: 100127. doi:10.1016/j.orp.2019.100127.

Pei Breivold, Hongyu. 2020. "Towards Factories of the Future: Migration of Industrial Legacy Automation Systems in the Cloud Computing and Internet-of-Things Context." *Enterprise Information Systems* 14 (4). Taylor & Francis: 542–62.

doi:10.1080/17517575.2018.1556814.

Porter, Michael E., and James E. Heppelmann. 2015. "How Smart, Connected Products Are Transforming Companies." *Harvard Business Review* 2015 (October). doi:10.1017/CBO9781107415324.004.

Provost, Foster, and Tom Fawcett. 2013. "Data Science and Its Relationship to Big Data and Data-Driven Decision Making." *Big Data* 1 (1). Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA : 51–59. doi:10.1089/big.2013.1508.

Ranjan, Shree, K Jha Jha, and Pralay Pal. 2016. "Antecedents and Imperatives of Integrating SCM and ERP in Manufacturing Organizations: A Coneptual Perspective with Big Data Analytics." *International Science Press*.

Saberi, Morteza, Omar Khadeer Hussain, and Elizabeth Chang. 2017. "Past, Present and Future of Contact Centers: A Literature Review." *Business Process Management Journal* 23 (3): 574–97. doi:http://dx.doi.org/10.1108/MRR-09-2015-0216.

Scholten, Bianca. 2007. *The Road to Integration: A Guide to Applying the ISA-95 Standard in Manufacturing*. International Society of Automation.

Sculley, D., Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean François Crespo, and Dan Dennison. 2015. "Hidden Technical Debt in Machine Learning Systems." *Advances in Neural Information Processing Systems* 2015-Janua: 2503–11.

Sein, K Maung, Ola Henfridsson, Sandeep Purao, Matti Rossi, and Rikard Lindgren. 2011. "Action Design Research." *MIS Quarterly* 35 (1): 37–56.

Shearer, Colin. 2000. "The CRISP-DM Model: The New Blueprint for Data Mining." *Journal of Data Warehousing* 5 (4): 23.

Stefanou, Constantinos J. 2001. "Organizational Key Success Factors for Implementing SCM/ERP Systems to Support Decision Making." *Journal of Decision Systems* 10 (1): 49–64. doi:10.3166/jds.10.49-64.

Subramaniyan, Mukund, Anders Skoogh, Hans Salomonsson, Pramod Bangalore, and Jon Bokrantz. 2018. "A Data-Driven Algorithm to Predict Throughput Bottlenecks in a Production System Based on Active Periods of the Machines." *Computers and Industrial Engineering* 125 (April). Elsevier: 533–44. doi:10.1016/j.cie.2018.04.024.

Trkman, Peter, Kevin Mccormack, Marcos Paulo Valadares De Oliveira, Marcelo Bronzo Ladeira, Marcos Paulo Valadares de Oliveira, and Marcelo Bronzo Ladeira. 2010. "The Impact of Business Analytics on Supply Chain Performance." *Decision Support Systems* 49 (3). Elsevier Science Publishers B. V.: 318–27. doi:10.1016/j.dss.2010.03.007.

Ugarte, B. Saenz De, A. Artiba, and R. Pellerin. 2009. "Manufacturing Execution System - A Literature Review." *Production Planning and Control* 20 (6): 525–39. doi:10.1080/09537280902938613.

Viaene, Stijn, and Annabel Van Den Bunder. 2011. "The Secrets to Managing Business Analytics Projects." *MIT Sloan Management Review* 53 (1): 65–69.

Wang, Meng, Mohammed Sadiq Altaf, Mohamed Al-Hussein, and Yongsheng Ma. 2018. "Framework for an IoT-Based Shop Floor Material Management System for Panelized Homebuilding." *International Journal of Construction Management* 0 (0). Taylor & Francis: 1–16. doi:10.1080/15623599.2018.1484554.

Weihrauch, Dennis, Paul Anton Schindler, and Wilfried Sihn. 2018. "A Conceptual Model for Developing a Smart Process Control System." *Procedia CIRP* 67. Elsevier B.V.: 386–91. doi:10.1016/j.procir.2017.12.230.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10): 1–23. doi:10.18637/jss.v059.i10.

Zhang, Xianyu and Ming, Xinguo, 2020. " Reference subsystems for Smart Manufacturing Collaborative System (SMCS) from multi-processes, multi-intersections and multi-operators" Enterprise Information Systems 14(3): 282-307 https://doi.org/10.1080/17517575.2019.1694705