Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 55-10 (2022) 2054–2059

Development of a collaborative tool for data valorisation in SMEs

Toumelin A.,* Agard B.,* Leduc M.**

* *Laboratoire en Intelligence des Données (LID)*
Département de mathématiques et génie industriel,
École Polytechnique de Montréal,
CP 6079, succursale Centre-Ville, Montréal, Québec, Canada
aime.toumelin@polymtl.ca, bruno.agard@polymtl.ca
** *Mon Système Fourrager*
maxime.leduc@gmail.com

Abstract: In recent years, solutions based on Industry 4.0. technologies have become more and more accessible for SMEs. As a result, better data collection strategies are being developed and an increase in available data has been observed. Furthermore, SMEs contribute largely to the GDP in certain industrial sectors such as the food industry, food services and the health sector. These SMEs are therefore naturally oriented towards Industry 4.0. solutions such as decentralized data collection and decision-making techniques. To an identical industrial sector, these SMEs can be considered a decentralized network of companies. There is interest in improving the overall performance of this type of network, while optimizing the individual performance of each company. In this paper, we propose a collaborative tool and a methodology to visualize and improve the performance of SME networks. The methodology is implemented through a case study in the agri-food sector in the Canadian forage industry. The collaborative platform that is developed enables a visualization of production performance, management techniques and the specific aspects of each individual company. Then, it proposes a methodology for the improvement of each company through best practices that are identified in similar contexts within the network.

Keywords: SME, decentralized network, Industry 4.0., forage industry.

1. INTRODUCTION

In Canada, 41.9% of the GDP is produced by SMEs (Government of Canada, 2020). In some sectors, SMEs are the main contributors, such as food services (64%), health care (86%), construction (65%), and agriculture (89%). In industrial sectors that are mainly composed of SMEs, the technological solutions that enable them to optimize, manage and influence the overall performance of production are limited (Stentoft et al., 2021). Moreover, these SMEs generally work in an isolated context and interact little with identical companies in their field (Luco et al., 2019). It is therefore difficult for these companies to estimate the quality of their production performance. The development of decentralized decision-making techniques and the emergence of industrialization 4.0 strategies allow those companies to envisage the emergence of new tools to improve how they operate (Agostini et al., 2015).Thus, the use of new types of sensors that are less expensive and more versatile empowers SMEs with innovative technological solutions that are better adapted to their needs. The development of new, more resilient data architecture based on cloud computing technologies encourages these SMEs to turn to data analysis tools. Furthermore, the increased availability of many open-source databases is expanding the data produced by these SMEs as well as producing additional knowledge useful to these companies. All of these solutions are promoting knowledge sharing and the popularization of decision support tools.Within this context, this article proposes a framework for the development of a collaborative platform aimed at improving the production performance of a network of SMEs.

Thus, section 2 presents the state of the art and in particular, the motivations and needs of such a tool (2.1), the specific context of the industrial sectors studied (2.2) as well as the existing technological solutions that are applicable to the context of SMEs (2.3). Section 3 presents the methodology used for the development of a collaborative platform. Section 4 applies the present methodology to a specific case study: the Canadian agri-food and forage production sector. To this end, the context of the business is first presented (4.1), the data preparation process is described (4.2) and the tool development methodology is outlined. Finally, sections 4.3 and 4.4 report the results as well as the limits of the platform developed, then provide some perspectives on how it can evolve in the future.

★ Canadian Forage and Grassland Association, and MITACS project IT23206

10.1016/j.ifacol.2022.10.010

2. STATE OF THE ART

2.1 Needs and Motivations

The decrease in sensor production costs, the advent of Cyber Physical Systems (CPS), the Internet of Things (IoT) and cloud computing technologies enable high-performance data collection systems to be built (Lee, 2017). These technologies are now implemented in the high-tech industry and are gradually being applied in SMEs (Mittal et al., 2018). However, the use of industrialization 4.0 technologies for data collection in small and medium-sized companies produces a considerable amount of data and contributes to the formation of decentralized networks (Mittal et al., 2018). A comparison of data from each company within such networks has not yet been achieved and therefore no knowledge can be extracted from it. In Canada, certain industrial sectors such as agri-food, manufacturing and construction are mostly represented by SME networks. The development of collaborative decision tools for these SMEs presents an intriguing reason to value such data. This article attempts to provide a collaborative tool that involves SMEs and allows for both the individual and global improvement of actors within this market.

2.2 The specific context of SMEs

Industrial sectors that are made-up of a majority of SMEs have some common characteristics. First of all, in such networks of companies, the geographical distribution of the actors is very wide (Villa, 2014). Moreover, the communication among each company is limited because of their large number and the competitive aspect of their activities (Moeuf et al., 2018). However, these companies produce the lion's share of the total production of the GDP. Each company collects its own data and the creation of a common initiative is difficult to achieve. Indeed, there are very few standards for data collection adapted to the context of SMEs (Stentoft et al., 2021). Secondly, the lack of means and knowledge limits the use of data when it is collected (Stentoft et al., 2021). The specificity of each industry represents a barrier to the pooling and valorization of these data. Finally, the data is rarely shared because it describes the company's performance and therefore must respect the anonymity and confidentiality of the data (Chonsawat and Sopadang, 2019). The development of a platform enabling the comparison of each individual company's performance, the identification of good practices, and the establishment of a productivity diagnosis while respecting the confidentiality and anonymity of each company's data fits perfectly into this SME context. The current solutions that exist and fit the needs of these companies in this context are presented in section 2.3.

2.3 Current technological solutions for SMEs

(Chonsawat and Sopadang, 2019) and (Mittal et al., 2018) developed methods for evaluating the maturation of 4.0 industrialization in the context of SMEs. These methods aim to evaluate the company's needs in terms of industrialization 4.0 and to estimate the added value of Industry 4.0. technologies to their business sector. Alcácer and Cruz-Machado (2019) provide an implementation framework for Industry 4.0. technologies and describe the main implementation bottlenecks for SMEs. Low financial resources, scarce use of advanced manufacturing technologies, and the specialization of products developed are the main factors that make the implementation of a solution from Industry 4.0 difficult for SMEs. Han and Trimi (2022) therefore propose a methodology for implementing Cyber Physical Systems (CPS) and Big Data to develop a generic data platform. This platform is adapted to the SME context and aims to provide a framework for data collection, data centralization and data valorization. Finally, Ren et al. (2015) develop theoretical models based on cloud manufacturing to increase communication between each SME in the network.

The use of collaborative platforms in an SME context brings about definite advantages by minimizing the costs associated with transportation and raw material acquisition. However, there is very little established with regards to a methodology that can build or solidify them. Indeed, the main obstacles to the concretization of such solutions come from operational realities related to management culture, behaviours, company knowledge and financial needs. Finally, the integration of these platforms raises challenges concerning interoperability, integration and automation as well as information diffusion management

2.4 Synthesis

The state of the art highlights the impact of industrialization 4.0 development on performance management for SME networks. On the one hand, the decrease in development costs and use costs of these technologies allows for the implementation of new engineering solutions adapted to the context of SME. On the other hand, the transition of Industry 4.0. technologies to SMEs for data collection techniques justifies the need for these SME networks to obtain data processing and valorization tools. Finally, if some technological solutions have been theorized in order to optimize the performance of such SME networks performances, there is a lack of concrete examples in scientific literature. Thus, this state of the art identifies a lack of collaborative tools that improve the performance of a SME's network. This research project tries to fill this gap by developing a collaborative tool intended for SMEs that allows them to analyze and improve their production performance. This article explains this process and shows its applicability through a case study on the development of a collaborative platform for the forage industry in Canada.

3. METHODOLOGY

The contribution of this research project is the development of a collaborative platform that improves the production performance of a network of SMEs. In order to reach this objective, a development methodology, inspired by the DMME method (Huber et al., 2019) and the CRISPP-DM method (Chapman et al., 2000) is adapted to the context of industrialization 4.0 and SMEs. It is composed of 5 steps and starts with (1) the acquisition of business expertise on the data application domain. The data collection and data preparation is then carried out in step 2. (3) The presentation of the individual performances is realized through a first visualization tool. Once the data is presented, step 4 improves the performance of each company. Finally, (5) the collaborative tool is deployed to end users. This development process is iterative, several feedback loops are realized between steps [2:3] and [1:4]. In order to accelerate the data development process and to adapt it to the needs of SMEs, the presentation step is used to identify some outliers in the data. Moreover, the performance improvement allows useful knowledge to be extracted for business insights. This knowledge is then reused to extract new data and to provide new indicators. The steps of the methodology are presented in Figure 1 and described in this section.

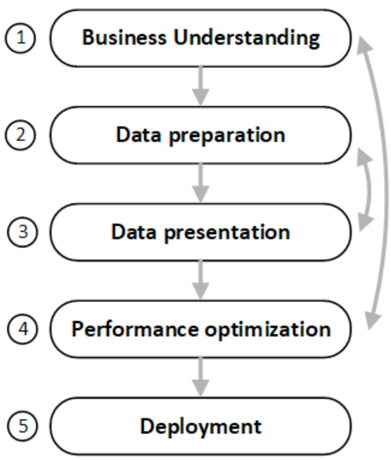

Fig. 1. Methodological framework

Business Understanding: The objectives of the "Business Understanding" step are to identify the needs of the domain and potential technological solutions that meet these needs. For SMEs, the databases are of a reasonable size. Moreover, these companies have very different architecture standards and the distribution of data sources varies greatly between each company.

Data preparation: The preparation step consists of the structuring and cleaning of the collected data. This step consists of identifying the type of data that is useful for building technological solutions, implementing collection solutions, building data pipelines and incorporating data cleaning strategies. During this step, data tables containing the list of platform users, the access they have to it, and the performance of each company must be built. Then, missing data in these tables is identified and corrected when possible. Otherwise, this data is deleted. Outlier detection is then performed based on the data preparation method from (Pyle, 1999). The main types of data present in the SMEs are temporal and quantitative production data. In addition, geospatial data as well as management practices data can also be exploited. Finally, each company has static data corresponding to its own characteristics.

Visualisation of current performances: Once the data is prepared, a data visualization tool is built to display the performance of each company. This step allows the data to first be validated by experts in the field. The presentation of the current performances provides a second opportunity to build key performance indicators specific to the application domain and provides a preliminary decision support tool for participating companies. In the context of SMEs, this step is divided into three distinct tasks: analysis of information flows, labelling of management practices and monitoring of production performance. Therefore, first dashboards allowing to monitor and describe the information flows produced by the company must be built. Then, all management practices must be collected and displayed. Each company then has a tool that allows it to understand all of the activities that impact its production performance. Lastly, a final table must be built to monitor the company's performance.

Individual performance improvement: Following the presentation of the basic performance of the companies, quantification, estimation and comparison modules are added in this step to optimize the individual performance of each company. First, tools for sharing good management practices based on existing theoretical and empirical models are implemented. Then, other modules for comparing the best practices present in the enterprise network are developed in order to classify the performance of each enterprise in groups according to its specific characteristics. At the end of this sub-step, clusters of companies with similar performance and characteristics are obtained. These similarities depend on the domain under study and can be based on size, geographical location and other attributes present in the database. The management practices of these groups can be analyzed and implemented through key performance indicators in the data platform. These performance indicators are ultimately shared for each company with similar characteristics.

Deployment: The deployment step takes place during the data presentation and after the improvement of the individual performance of each company. During this step, particular attention is paid to preserving the anonymity of the partners. Also, the confidentiality of the data used must be respected. Finally, key performance indicators developed during the performance improvement step must guarantee the security and integrity of the data during deployment.

Following the steps described above, companies have a collaborative tool that is capable of optimizing their production performance. Finally, other perspectives can be raised based on this methodology. The next steps aim at improving the overall performance of the enterprise network by using machine learning tools and will be developed in further research.

4. CASE STUDY

4.1 Business understanding: the Canadian forage industry

The forage industry is an industrial sector with the objective of producing forage for animal consumption. In Canada, the forage industry is mainly represented by the production of alfalfa crops. This plant is a leguminous that is grown for its high protein content. Specificity of alfalfa crops relies on its persistence. Alfalfa is harvested from 3 to 5 times a year for an average exploitation time of 4 years. In Canada, 89.5% of the total production is realized by small producers (Government of Canada, 2020). These farms are spread throughout the country, with a higher concentration in Manitoba, Saskatchewan and Quebec. All of these companies form a decentralized network where each farm faces very different environmental parameters. 240 farms were surveyed as part of this project. In order to collect a maximum amount of information on the performance of these farms, multiple sensors are installed and will collect data for the needs of the project for 3 years (2021-2023). The work described in this article is based on the data collected in the first year (2021).

4.2 Data preparation

The data collected in this project is meteorological data, data related to crop management practices, agronomic data and production yield data. The farms surveyed are distributed over a wide area, so an adapted data collection strategy had to be used. The weather data was acquired through the Meteostat python API, which captures daily temperature, pressure and precipitation from a worldwide network of weather stations. To ensure accuracy, all selected stations must be located within a 50 km radius of the farm that is being studied. Also, for the purposes of the project, a data collection application (CC+) was developed and made available to the companies. It allows them to enter data on crop management (harvests, organic and mineral fertilization) and production yields. Finally, the soil analysis data comes from a database made available by an agronomic consulting partner involved in the project. The following table lists all of the families of parameters collected for this project. Obviously, each family of parameters groups together several collected attributes.

Table 1. Collected parameters

Type	Parameters	Resolution	Source
Weather	Temperature	Daily	API
	Precipitations	Daily	API
	Pressure	Daily	API
Agronomic	Forage yield	Per cut	CC+
	Harvest date	2-5 times/year	CC+
	Stem count	2 times/year	CC+
	Organic Fertilisation	2-5 times/year	CC+
	Mineral Fertilisation	Yearly	CC+
	Forage quality	Per cut	CC+
Soil	Mineral quantity (P,K,S,N)	Every 5 years	Partner

Data cleaning was done in several steps. First of all, the collected data sources are made up of manually collected tables. Several types of errors are then detected. Hence, the null values are identified and then filled based on (Kim et al., 2003) and (Gschwandtner et al., 2012). When not applicable, each company is interviewed to recover missing values. Following this step, a report is written to improve the data collection methodology for future years. Next, outliers are identified by selecting all values greater than 2 times the standard deviation. These attributes are analyzed and 2 types of errors are corrected. Thus, unit errors are identified, converted to the International system and the typographical errors are corrected when possible or collected again from the producers. These previous steps allow us to obtain first validity intervals for each attribute. Then, these intervals are put into perspective with the business understanding steps and validated by industry experts according to their geographical context and field of study. Robust filters are then confirmed and allow the processing and validation of the database. These filters finalize the data cleaning step and result in a clean database. The next step uses this database to display the performance parameters of each company and produces useful insights for producers.

4.3 Visualisation of current performance

The current performance of each farm is displayed on dashboards using the Tableau platform in figures 2, 3 and 4. Three visualization tools are built: a timeline, a performance table and a map presenting each field characteristics.

Timeline:

The timeline tool displays different management practices of the farmers. Thus, during a farming season, the farmer performs mineral fertilization practices, organic fertilization practices, seeding and harvesting steps. Each operation correlated with the climatic component has an influence on yields and production quality. The date, quantity and type of these operations are therefore collected and displayed in this first visualization tool. This data can then be compared by region to highlight good and bad practices. This first tool is presented in Figure 2.

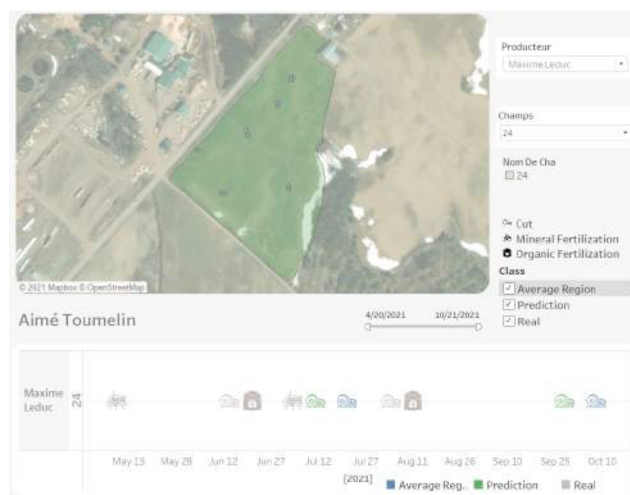

Fig. 2. Management practices evaluation

The next steps in the development of this tool are to display the average of each operation in relation to similar farms. Then, theoretical and predictive tools can be used to advise the producer on the best time to perform each management operation in these fields.

Production yield:

The second tool integrated into the platform highlights the production performance of each field on the farms. In this regard, the average yield per field is displayed over years of operation. In addition, the total cumulative yield per farm is also presented in order to obtain an idea of the overall performance of the farm. Finally, the yield per cut for each field is analyzed to show the evolution of field performance over time and to study the persistence of each alfalfa field. The yields of each farm are then compared to the geographic average of the 5 closest farms. This tool is presented in Figure 3.

Fig. 3. Yield visualization dashboards

Possible improvements to this tool are the prediction of yields for future cuts as well as production quality based on previous forage analysis.

Health and field characteristics

The third tool was developed to study the health of the cultivated fields. The data used to analyze the soil fertility parameters were taken from a soil analysis of each field that was done at least every 5 years as well as from the forage analysis. This data is displayed on a map that provides access to a soil analysis of the field and an estimation of the health of the field. Thus, colourful graphic indicators are displayed to synthesize the results of these analyses. The colors of each field indicate the overall health of the field. This tool is presented in figure 4.

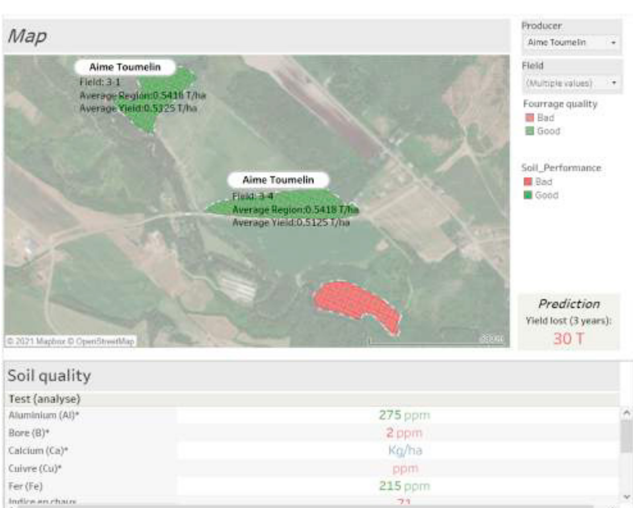

Fig. 4. Health and field characteristics

The tools presented above offer examples of performance indicators that can illustrate the individual performance of each partner. These indicators have been determined according to the context of application and can be used in other fields of study. For example, a similar method could be applied to compare individual SME data with the respective network data. These examples therefore illustrate the feasibility of this method without constraining the confidentiality and security of each company's data. The next step in the development of this tool is the use of machine learning techniques to identify and determine the environmental parameters that influence the degradation of the characteristics of each field as well as potential yield lost if any correction is made.

4.4 Individual performance improvement:

Feature engineering techniques are performed to extract knowledge from the database prepared. Based on domain expertise, agronomic parameters are added such as GDD (Growing Degree Day), cutting management and general soil fertility and health. In addition, the use of supervised learning and segmentation techniques allows the creation of clusters and the classification of their performance into categories that can be interpreted by producers and agronomists. These classifications are described and displayed directly to producers through the diagnostic tool. A view of this classification is shown in Figure 5.

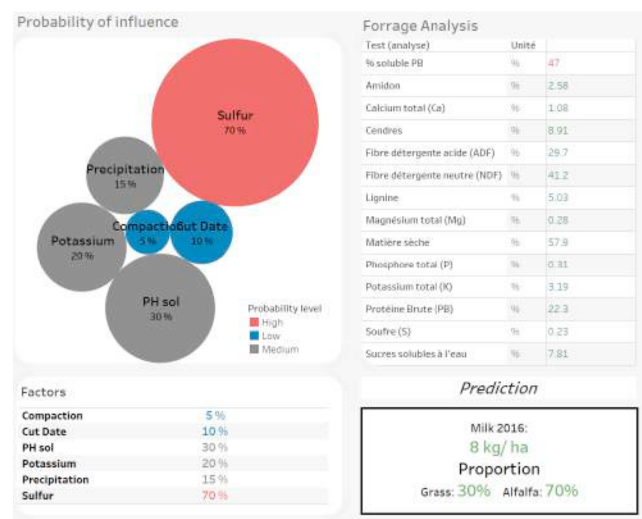

Fig. 5. Parameter improvement

Figure 5 identifies the factors that have the most negative impact on the farm. In addition, a forage quality analysis is interpreted and presented at the top right of the figure. This view supports an agronomists' analysis of the performance and health of the field being studied.

4.5 Deployment:

Finally, as described in section 2, SMEs do not have a lot of financial resources. As a result, it is necessary to quickly provide a tool that brings value to SMEs. In this regard, the deployment of this tool is realized iteratively following the methodology presented in section 3. Figure 6 presents the architecture of the developed tool.

As presented in figure 6, the collected data is gathered in three data sources. The first data source contains the weather data of each farm. It is hosted on a Hobolink platform and a data pipeline is built to link it directly to

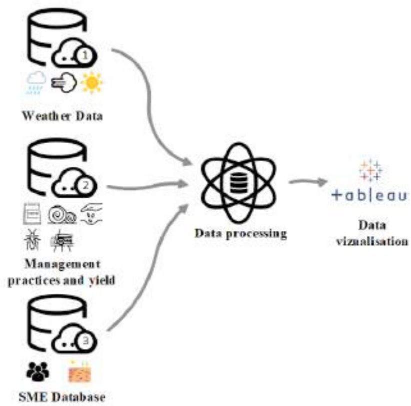

Fig. 6. Architecture of the datasources

the python script to prepare the data. Database 2 hosts data on yield and management practices and is built on the Microsoft JIRA solution. Finally, database 3 comes from an agronomic partner that has the soil analysis and producer data for each farm.

5. CONCLUSION

The goal of this research project was to develop collaborative tools that improve the individual production performance of companies within an SME network. The state of the art carried out in section 2 highlighted the cost benefits of entering into and using Industry 4.0 tools, allowing the transition of these technologies towards the SMEs. In particular, it has been demonstrated that these SMEs already use data collection techniques from Industry 4.0. There is, therefore, a need for these SMEs to create tools for subsequent stages of data collection such as data processing and data valorization. Although some performance improvement tools have already been theorized in the literature, few collaborative tools have been fully developed and implemented for SMEs. The objective of this paper was to present a methodology for developing collaborative tools adapted to SMEs. In order to demonstrate a practical approach of this methodology as well as its success, it was applied to a case study of the agri-food industry in Canada (more precisely, the forage industry). Tools for presenting and improving the performance of these SMEs were developed and implemented using this methodology. The deployment of these tools has allowed other shortcomings related to this issue to be identified. Thus, perspectives of evolution could be listed and the next objectives of this research topic will aim towards the overall improvement of the enterprise network. Then, it would be relevant to improve this tool in order to communicate and influence good practices within the business networks. An analysis of the best performing companies, their practices as well as the least performing companies of the network, could help optimize the performance of SMEs. A comparison of the performance of each company with its neighbors could be developed to facilitate the implementation of new measures and to limit resistance to operational changes. Finally, identifying the factors that have the most impact on decreases in performance could enable the the impact of these factors to be quantified, and to improve the performance of the overall SME network.

REFERENCES

Agostini, L., Filippini, R., and Nosella, A. (2015). Management and performance of strategic multipartner sme networks. *International Journal of Production Economics*, 169, 376–390.

Alcácer, V. and Cruz-Machado, V. (2019). Scanning the industry 4.0: A literature review on technologies for manufacturing systems. *Engineering Science and Technology, an International Journal*, 22(3), 899–919.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9.

Chonsawat, N. and Sopadang, A. (2019). The development of the maturity model to evaluate the smart smes 4.0 readiness. In *Proceedings of the international conference on industrial engineering and operations management*.

Government of Canada (2020). Key small business statistics — 2020. Technical report.

Gschwandtner, T., Gärtner, J., Aigner, W., and Miksch, S. (2012). A taxonomy of dirty time-oriented data. In *International Conference on Availability, Reliability, and Security*, 58–72. Springer.

Han, H. and Trimi, S. (2022). Towards a data science platform for improving sme collaboration through industry 4.0 technologies. *Technological Forecasting and Social Change*, 174, 121242.

Huber, S., Wiemer, H., Schneider, D., and Ihlenfeldt, S. (2019). Dmme: Data mining methodology for engineering applications – a holistic extension to the crisp-dm model. *Procedia CIRP*, 79, 403–408. 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy.

Kim, W., Choi, B.J., Hong, E.K., Kim, S.K., and Lee, D. (2003). A taxonomy of dirty data. *Data mining and knowledge discovery*, 7(1), 81–99.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303.

Luco, J., Mestre, S., Henry, L., Tamayo, S., and F.Fontane (2019). Industry 4.0 in smes: A sectorial analysis. *IFIP Advances in Information and Communication Technology,Springer, Cham.*, vol 566.

Mittal, S., Khan, M.A., Romero, D., and Wuest, T. (2018). A critical review of smart manufacturing & industry 4.0 maturity models: Implications for small and medium-sized enterprises (smes). *Journal of Manufacturing Systems*, 49, 194–214.

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., and Barbaray, R. (2018). The industrial management of smes in the era of industry 4.0. *International Journal of Production Research*, 56(3), 1118–1136.

Pyle, D. (1999). *Data preparation for data mining*. Morgan Kaufmann.

Ren, L., Zhang, L., Tao, F., Zhao, C., Chai, X., and Zhao, X. (2015). Cloud manufacturing: from concept to practice. *Enterprise Information Systems*, 9(2).

Stentoft, J., Wickstrøm, K.A., Philipsen, K., and Haug, A. (2021). Drivers and barriers for industry 4.0 readiness and practice: empirical evidence from small and medium-sized manufacturers. *Production Planning & Control*, 32(10).

Villa, A. (2014). *Managing cooperation in supply network structures and small or medium-sized enterprises*. Springer.