An Industry 4.0 Intelligent Decision Support System for Analytical Laboratories

António João Silva and Paulo Cortez(✉)

ALGORITMI Center, Department of Information Systems,
University of Minho, 4804-533 Guimarães, Portugal
id7322@alunos.uminho.pt, pcortez@dsi.uminho.pt

Abstract. This paper presents an Intelligent Decision Support System (IDSS) to enhance the management of Analytical Laboratories (AL) of a company operating in the chemical industry. This IDSS incorporates two predictive Machine Learning (ML) models, related with the prediction of the arrival of samples at the AL and the consumption of AL materials, which are then used to perform prescriptive analytics for AL instrument allocation tasks. The IDSS is also complemented with descriptive analytics of instrument similarities regarding the tests performed for better supporting the AL manager decisions. The IDSS includes interactive dashboards and it was successfully validated by the AL managers using the Technology Acceptance Model (TAM) 3 and open interviews, which resulted in a positive feedback.

Keywords: Intelligent Decision Support System · Dashboards · Machine Learning · Industry 4.0

1 Introduction

Industry 4.0 has recently emerged and with it there has been an increasing amount of digitalized data that reflects industrial processes. Within this context, Intelligent Decision Support Systems (IDSS) [2] can be very useful to extract valuable insights from the industrial data, allowing to enhance several business processes. An IDSS is a decision support system that uses Artificial Intelligence techniques (e.g., Machine Learning, Metaheuristics) to enhance managerial decisions [13]. In this work, we propose an IDSS that is based on descriptive, predictive and prescriptive analytics, aiming to assist the managerial decisions of Analytical Laboratories (AL) from a Chemical Industry that is being transformed through the Industry 4.0 concept.

In previous works, we have proposed Machine Learning (ML) solutions to assist some partial AL tasks: predict the arrival time of In-Process Control (IPC) samples at the quality testing laboratories [18]; and estimate the AL materials consumption based on weekly plans of AL sample analyses [17]. In this paper, we

Published by Springer Nature Switzerland AG 2022
I. Maglogiannis et al. (Eds.): AIAI 2022, IFIP AICT 647, pp. 159–169, 2022.
https://doi.org/10.1007/978-3-031-08337-2_14

present the full IDSS that integrates both predictive analytics, supporting the allocation of AL instruments (prescriptive analytics). The IDSS is also complemented with descriptive analytics executed over AL historical records, allowing the AL managers to better identify similarities among instruments. Prior to the Industry 4.0 transformation, the relevant digital records were spread in distinct databases, located in different departments (production and the AL), making the AL manager decisions more difficult. The proposed IDSS integrates all relevant data records into a single data repository, while also providing the business analytics results in terms of an interactive visual tool based on dashboards. A IDSS prototype was deployed in the chemical company and then evaluated by the AL managers by answering a questionnaire built using the Technology Acceptance Model (TAM) 3 model [22] and by using open interviews.

2 Related Work

Within the Industry 4.0 concept, there are several studies proposing data-based interactive dashboards. For instance, our survey about the usage of Business Analytics in Industry 4.0 [19] has found several examples of dashboards used to monitoring the production process, as well as verify new insights on the shop floor [14,15]. Moreover, in the automotive industry, data-based dashboards were used to monitor the assembly processes [20]. Also in the manufacturing sector, sensors and Internet-of-Things (IoT) data were also integrated into dashboards to monitor the productive process [12]. Concerning the specific chemical industry, we have found one one dashboard example that was proposed to control and monitor the production of a chemical plant [3].

Turning to the incorporation of Artificial Intelligence techniques for decision support, there are a few studies that integrate ML results in dashboards. For instance, a few examples are: use Neural Networks to improve the energy saving in factories [11]; usage of a Random Forest algorithm and IoT sensors to improve fault diagnosis tasks [21]; and a predictive maintenance system using a Remaining Useful Life (RUL) model to estimate the health index of production machines [5]. However, regarding the application of IDSS in the AL of chemical industry the research is very scarce. This occurs because the AL are mostly managed manually, where Information Technology (IT) is mostly focused on storing the quality values and not the AL processes. Following an Industry 4.0 process transformation, we have previously developed two ML works, aiming to empower the AL of a chemical company with two data-driven models: to predict the arrival of In-Process Control (IPC) samples at the ALs [18]; and to predict the weekly consumption of AL materials [17]. In this paper, we present the full IDSS that provides interactive dashboards that integrate these two ML models and also descriptive analytics (for instruments allocation and similarities).

3 Materials and Methods

3.1 Problem Formulation

The analyzed company is from the Chemical Sector and it includes three main areas: Warehouse, Production and Analytical Laboratories (AL). The Warehouse is where the raw materials are received. It is also the destination of the products produced before being shipped to the customers. The Production area is where the chemical products are manufactured. Finally, the AL are responsible for testing all products and raw materials, checking if they meet the required quality standards. Before adopting an Industry 4.0 transformation, the entire communication process between these three areas was mainly manual and there was no real-time monitoring of the industrial processes, often leading to delays in the preparation of production materials or in the analyzes performed by the AL. These delays strongly affected deadlines for production plans.

Concerning the AL, these involve human analysts, instruments and several types of samples, namely raw materials, In-Process Control (IPC) and Final Products, that need to be analyzed, i.e., allocated into one or more analytical instruments. In particular, In Production samples are a priority because if they are not analyzed in a timely manner, the production process may stop. Each instrument allocation requires time and manual effort, to prepare and conduct the analysis and then collect the obtained results. There is an information system that records all quality test data, but such IT is mostly focused on the testing measurements and not on the AL processes. Thus, the management of the AL (e.g., human resource and instrument allocation planning, sample prioritization, prior preparation of instruments), assumes a strong manual effort, which is difficult due to the lack of a real-time data communication with the Warehouse and Production areas.

3.2 Proposed IDSS

To solve the previous mentioned AL management issues, and benefiting from the Industry 4.0 transformation performed at the company, we propose an IDSS that incorporates descriptive, predictive and prescriptive analytics. The proposed IDSS architecture is depicted in Fig. 1. It includes two main layers. The Big Data layer is responsible for extracting and processing data from the different databases used in the organization. Indeed, the IDSS consumes the data from the different areas and applications from the organization (e.g., Warehouse, Production, AL), resulting as the ground truth data repository for the AL. The processed data is then fed into the Data Analytics layer, which incorporates descriptive, predictive and prescriptive analytics for AL management.

The developed tool includes two predictive models that were previously studied. Both models are based on an Automated ML (AutoML) procedure but fed with different input attributes and training data. The adopted AutoML H2O tool (https://www.h2o.ai/) automatically selects the best regression model among 6 families of algorithms: Random Forest (RF), Extremely Randomized Trees (XRT), Generalized Linear Model (GLM), GBM, XGBoost (XG) and a Stacked

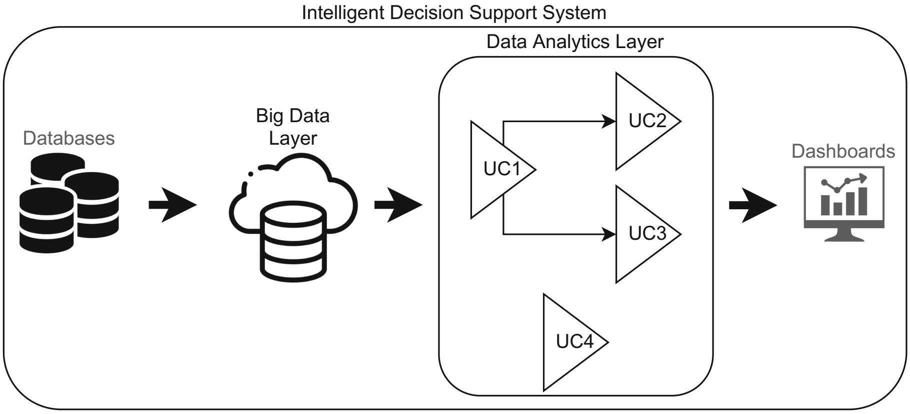

Fig. 1. Proposed Architecture

Ensemble (SE). The proposed IDSS includes an extension of the first predictive model, termed here Use Case (UC) 1 (UC1), successfully tested for estimating the arrival of IPC samples at the ALs [18]. In ths proposed IDSS, the model is adapted to perform predictions for all types of AL samples (the studied IPC and also the raw materials and final products). It should be noted that each sample arrived at the AL is associated with a fixed set of quality tests to be executed. The IDSS also integrates a second predictive model (UC2) that estimates the weekly consumption of AL materials [17]. This second predictive model requires, as input, a weekly plan of quality tests to be performed, which is built in advance by adopting the UC1 predictive model. The IDSS also includes prescriptive analytics (UC3), which is based on sample arrival estimates (UC1) and historical records regarding previous instrument allocations, allowing to provide suggestions of future instrument allocation. Finally, the IDSS also includes descriptive analytics set in terms of historical associations of instruments to quality tests (UC4), allowing to identify instrument similarities. All analytics are incorporated into friendly user dashboards.

Regarding the UC3, to issue recommendations of AL instruments allocation, we use a statistical approach that considers the UC1 predictions (tests to be executed) and that are matched with historical records of instrument allocation. For each required test, we assume as the "best" analytical instrument, the one currently available that has been mostly used for executing such test. An instrument is considered available if the its scheduled weekly allocation is lower than 70% (a value that was defined by the AL experts). Once an instrument is allocated, the IDSS is refreshed, with the allocation records being updated.

Finally, the UC4 is based on an $I \times T$ matrix computed using historical records and that measures the total number of tests ($t \in T$) executed by an instrument ($i \in I$). Then, the known Pearson correlation is used to compute the association between two rows of the matrix (i.e., two instruments). In our dashboards, the correlation matrix [8] is shown as a colored heatmap, where more similar instruments are signaled by a stronger red color.

3.3 Evaluation

The proposed IDSS was developed by a research team that included both Artificial Intelligence and Chemical company experts but not the direct AL managers. Thus, to properly evaluate the IDSS, we adopted the Technology Acceptance Model (TAM) 3 [22], allowing to define a questionnaire that contains 10 questions and that was answered by the AL managers after experimenting the proposed tool. The questionnaire assumes the following TAM 3 constructs: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Perception of External Control (PEC), Job Relevance (REL), Output Quality (OUT), and Behavioral Intention (BI). Each question included a 5-point likert scale option for each answer, ranging from 1 (extremely disagree) to 5 (extremely agree). These questionnaires were complemented by a direct feedback from the AL managers, obtained by using open interviews in which the manager freely provided their opinions about the proposed IDSS. Furthermore, we also map the capabilities of the proposed IDSS tool, which are compared with the currently available AL informational processes (denoted as "As-Is") [7].

4 Results

4.1 Developed IDSS Prototype

The designed IDSS was written using the R language, with the ML solutions being developed using specific R [16] packages, namely `rminer` [6], `H2O AutoML` [1], `Forecast` [9,10] and `shiny` [4]. The IDSS was fed with real-world data from the analyzed chemical company, collected from January 2016 to May 2019 and that results from a merge of the different databases adopted by the organization.

The user interface was developed using `shiny` and it includes three main dashboards to present the descriptive (UC4), predictive (UC1 and UC2) and prescriptive (UC3) analytics. The first dashboard presents: the expected arrival of samples and quality tests to be carried out in the current week (UC1); the expected raw material consumption (UC2); the history of quality analyzes carried out in the previous week; and an overview of the historical arrival of samples to the laboratory in the last year. The second dashboard shows the current allocation of AL instruments and suggestions on the best instrument to be used for each planned test (UC3). Finally, the last dashboard contains the correlation heatmaps based on the $I \times T$ association matrix (UC4).

The first dashboard is presented in Fig. 2 and it contains three components. The first one is the top bar that shows warnings about issues that could occur during the current week. This includes information about how many instruments have an expected occupation above 50%, the number of analyzes without any instrument usage history, as well as the progress of test analyzes for the current day (in Fig. 2, this value is set at 0%). The second middle component includes three tables, presenting: the daily sample arrival (UC1) predictions (left table); how many analyzes are planned to be carried out on the current day (middle table); and the predicted weekly AL material consumption (UC2, right table).

The third bottom component has two graphs. The first plot (bottom left) shows the number of samples that arrived at the laboratories every week by type (IPC, Raw Material, Final Product), while the second graph (bottom right) displays the number of analyzes performed per week by sample type.

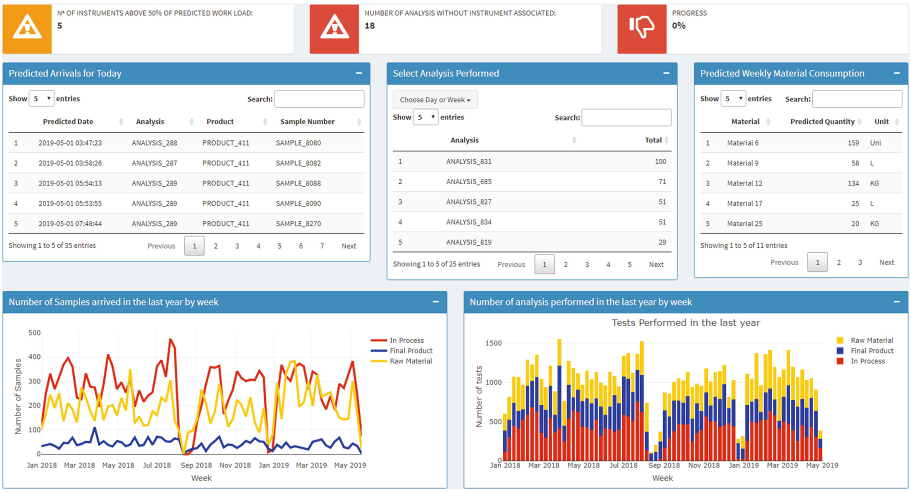

Fig. 2. Example of the first IDSS dashboard

The selection of the IDSS top menu tab allows the access to the second dashboard (Fig. 3). The top left component "Analysis to be performed in this week" allows to select a quality test, refreshing the middle barplot graphs that show the instruments that are used for that specific test and sample (left) or just for that specific test (without sample specification, right plot). At the same time, the table on the top right presents the UC3 results as the suggested instrument to be assigned to that specific test analysis, along with the load work for the same instruments for that week. Finally, the bottom left table contains the information about the tests that have no historical records of instrument usage.

The last and third dashboard is presented in two figures and it is related with the UC4 descriptive analytics. Figure 4 displays the correlation tables for a given instrument divided by two groups of instrument machines: HPLC (left table) and GC (right table). The top buttons ("Chosse HPLC/GC") allows the user to select one instrument from the displayed list. Once the instrument is selected, a table is displayed, sorting in a descending order the correlation values of most similar instruments. The third column on the tables shows the most used test analysis for each instrument. The bottom part of the third dashboard is presented in Fig. 5, which shows the instrument correlation heatmaps for each group of instruments. The heatmap provides easy visualization of the most correlated HPLC and GC instruments.

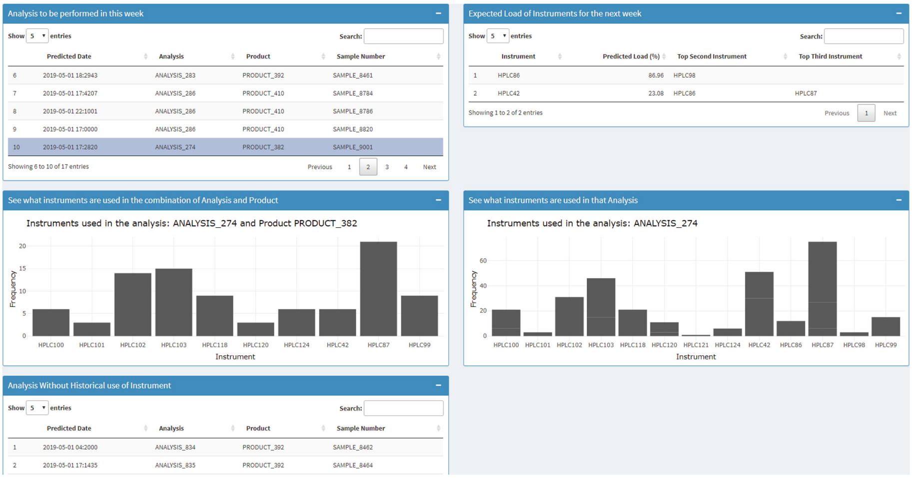

Fig. 3. Example of the second dashboard

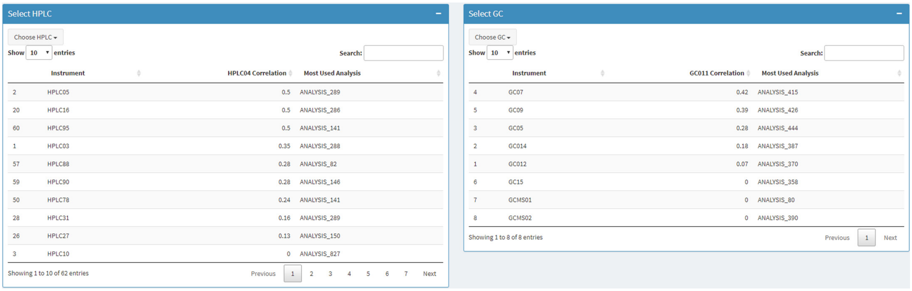

Fig. 4. Example of the third dashboard (instruments correlation)

4.2 Evaluation

The designed TAM 3 questionnaire is shown in Table 1. The obtained results are presented in Table 2, where each value corresponds to the average of two laboratory managers. We note that these managers correspond to IT AL staff from the analyzed chemical company and that were not directly involved in the presented research. The average responses are between 3.5 (70%) and 4 (80%), which means that laboratory managers had a positive acceptance of our IDSS. The most positive answers were related with the Perceived Usefulness (PU1 and PU2), Job Relevance (Rel 2) and Behavioral Intention (BI1). After obtaining the questionnaire responses, we have performed individual interviews, where the AL managers provided more specific feedback about the proposed IDSS. Regarding the first IDSS dashboard, both managers agreed that the information provided was simple and objective, being valuable to help the analysts to prepare the materials and the laboratory before the sample arrival. Turning to the second

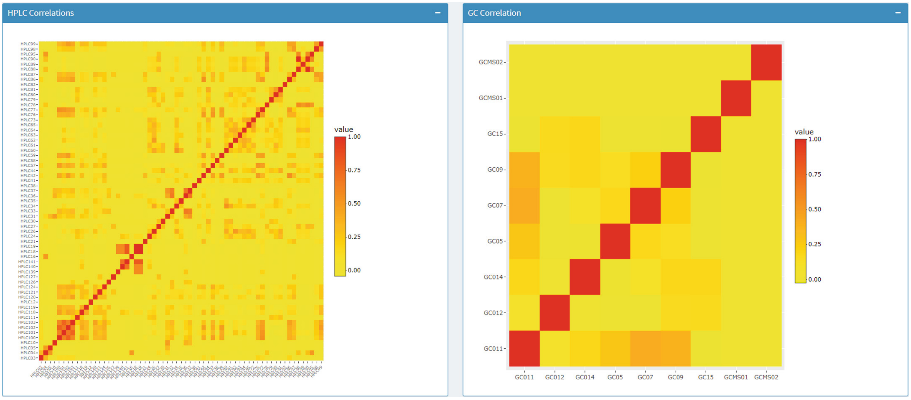

Fig. 5. Example of the third dashboard (instruments heatmap)

Table 1. The adopted TAM 3 questionnaire

Construct	Items	Question
Perceived Usefulness (PU)	PU1	Using the Dashboards improves my performance in my job
	PU2	The Dashboards are (potentially) useful in my job
Perceived Ease of Use (PEOU)	PEOU1	I find the Dashboard interface to be easy to use
	PEOU2	It's easy to get the information that I want from the Dashboards
Perceptions of External Control (PEC)	PEC1	I have the knowledge to use the Dashboards
Job Relevance (REL)	REL1	In my job, the usage of the Dashboards is important
	REL2	The use of the Dashboards is pertinent to my various job-related tasks
Output Quality (OUT)	OUT1	The quality of the output I get from the Dashboards is high
	OUT2	I have no difficulty telling others about the results ofusing the Dashboards
Behavioral Intention (BI)	BI1	Assuming I had access to the Dashboard, I intend to use it

Table 2. The TAM 3 questionnaire results (average of two responses).

PU1	PU2	PEOU1	PEOU2	PEC1	REL1	REL2	OUT1	OUT2	BI1
4	4	3.5	3.5	3.5	3.5	4	3.5	3.5	4

dashboard, related with the instruments load, they found it interesting but signaled the lack of information about new instruments and analyses. As for the third dashboard, it was considered helpful, particularly the correlation heatmap, which can be useful to identify new groups of instruments. However, such identification needs to be complemented by human domain knowledge, since there are instruments within the same group that can have different capabilities (e.g., refractive-index or infra-red). The AL managers also considered the dashboard useful to check if there a overlap between groups of instruments and if new groups of instruments could be defined. Overall, the AL managers concluded that the proposed IDSS (including its three dashboards), is valuable for planning the analyzes to be carried out on the samples, to improve the instrument allocation and to know how many analyzes will be carried out. Table 3 summarizes the main features introduced by the proposed IDSS, which substantially enhance the capabilities currently available at the AL (As-Is).

Table 3. Comparison between the current AL (As-Is) and proposed IDSS informational processes.

Capabilities	As-Is	IDSS
Historical overview of samples arrived	✓	✓
Historical overview of analysis performed	✓	✓
Sample arrival prediction		✓ (UC1)
Weekly estimates of materials consumption		✓ (UC2)
Expected instruments load		✓ (UC3)
Suggested allocation of instruments		✓ (UC3)
Information of analysis without instruments	✓	✓
Visualization of instrument similarities		✓ (UC4)

5 Conclusions

In this paper, we present an Intelligent Decision Support System (IDSS) that was developed for the Analytical Laboratories (AL) of a chemical company that is being transformed under the Industry 4.0 concept. The proposed IDSS includes two main layers: Big Data – responsible for extracting and processing data from different data sources, leading to a single and updated AL data repository; and Data Analytics – which includes descriptive, predictive and prescriptive analytics that aim to enhance the managerial decisions performed by AL managers.

Using recent data from a real-world chemical company, in previous works we have proposed two predictive analytics (IPC sample arrival prediction – UC1 and weekly AL materials consumption – UC2). The Data Analytics layer includes these analytics, extending the arrival prediction capabilities to all AL sample types (e.g., raw materials and final products). Moreover, it includes a novel prescriptive method (UC3) for suggesting instrument allocations for quality tests based on historical records and the sample arrival predictions (UC1). Finally, it includes descriptive analytics regarding laboratory instrument similarities (UC4). A IDSS prototype was developed, which integrated all proposed analytics in three main interactive dashboards and used data collected from January 2016 to May 2019. The prototype was evaluated by two AL managers that were not directly involved in the IDSS design by adopting Technology Acceptance Model (TAM) 3 questionnaires and open interviews. Overall, a very positive feedback was obtained. In particular, the proposed IDSS was considered valuable to better prepare and assign instruments to samples, as well as to better estimate the ammount of quality tests that will be carried out.

In future work, we intent to add new modules to the IDSS that are oriented to the maintenance of the instruments. The goal is to predict corrective maintenance actions and also support the scheduling of preventive maintenance operations.

Acknowledgments. This work has been supported by FCT – Fundação para a Ciência e a Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. The authors also wish to thank the chemical company staff involved with this project for providing the data and also the valuable domain feedback.

References

1. Aiello, S., Eckstrand, E., Fu, A., Landry, M., Aboyoun, P.: Machine learning with r and h2o. In: H2O Booklet (2016)
2. Arnott, D., Pervan, G.: A critical analysis of decision support systems research revisited: the rise of design science. J. Inf. Technol. **29**(4), 269–293 (2014). https://doi.org/10.1057/jit.2014.16
3. Bellini, P., Cenni, D., Mitolo, N., Nesi, P., Pantaleo, G., Soderi, M.: High level control of chemical plant by industry 4.0 solutions. J. Indust. Inf. Integrat. **26**, 100276 (2022). https://doi.org/10.1016/j.jii.2021.100276
4. Chang, W., et al.: Shiny: Web Application Framework for R. r package version 1.7.1 (2021). https://CRAN.R-project.org/package=shiny
5. Chiu, Y.C., Cheng, F.T., Huang, H.C.: Developing a factory-wide intelligent predictive maintenance system based on industry 4.0. J. Chinese Inst. Eng. **40**(7), 562–571 (2017). https://doi.org/10.1080/02533839.2017.1362357
6. Cortez, P.: Modern Optimization with R. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08263-9
7. Darwish, A.: Business Process Mapping: A Guide to Best Practice. Writescope Publishers (2011)
8. Ferré, J.: 3.02 - regression diagnostics. In: Brown, S.D., Tauler, R., Walczak, B. (eds.) Comprehensive Chemometrics, pp. 33–89. Elsevier, Oxford (2009). https://doi.org/10.1016/B978-044452701-1.00076-4
9. Hyndman, R., et al.: Forecast: Forecasting Functions for Time Series and Linear Models. r package version 8.13 (2020). https://pkg.robjhyndman.com/forecast/
10. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. J. Statist. Softw. **26**(3), 1–22 (2008)
11. Kabugo, J.C., Jämsä-Jounela, S.L., Schiemann, R., Binder, C.: Industry 4.0 based process data analytics platform: a waste-to-energy plant case study. Int. J. Electric. Power Energy Syst. **115**, 105508 (2020). https://doi.org/10.1016/j.ijepes.2019.105508
12. Mahmoodpour, M., Lobov, A., Lanz, M., Mäkelä, P., Rundas, N.: Role-based visualization of industrial iot-based systems. In: 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), pp. 1–8 (2018). https://doi.org/10.1109/MESA.2018.8449183
13. Michalewicz, Z., Michalewicz, M.: Machine intelligence, adaptive business intelligence, and natural intelligence [research frontier]. IEEE Comput. Intell. Mag. **3**(1), 54–63 (2008). https://doi.org/10.1109/MCI.2007.913389

14. Neuböck, T., Schrefl, M.: Modelling knowledge about data analysis processes in manufacturing. IFAC-PapersOnLine **48**(3), 277–282 (2015). https://doi.org/10.1016/j.ifacol.2015.06.094
15. Niño, M., Blanco, J.M., Illarramendi, A.: Business understanding, challenges and issues of big data analytics for the servitization of a capital equipment manufacturer. In: IEEE International Conference on Big Data (Big Data), pp. 1368–1377 (2015). https://doi.org/10.1109/BigData.2015.7363897
16. R Development Core Team. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria (2008). http://www.R-project.org, ISBN 3-900051-07-0
17. Silva, A.J., Cortez, P.: An automated machine learning approach for predicting chemical laboratory material consumption. In: Maglogiannis, I., Macintyre, J., Iliadis, L. (eds.) AIAI 2021. IAICT, vol. 627, pp. 105–116. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79150-6_9
18. Silva, A.J., Cortez, P., Pilastri, A.: Chemical laboratories 4.0: a two-stage machine learning system for predicting the arrival of samples. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds.) AIAI 2020. IAICT, vol. 584, pp. 232–243. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49186-4_20
19. Silva, A.J., Cortez, P., Pereira, C., Pilastri, A.: Business analytics in industry 4.0: a systematic review. Exp. Syst. **38**(7), e12741 (2021). https://doi.org/10.1111/exsy.12741
20. Silva, N., et al.: Advancing logistics 4.0 with the implementation of a big data warehouse: a demonstration case for the automotive industry. Electronics **10**(18) (2021). https://doi.org/10.3390/electronics10182221
21. Tran, M.Q., Elsisi, M., Mahmoud, K., Liu, M.K., Lehtonen, M., Darwish, M.M.F.: Experimental setup for online fault diagnosis of induction machines via promising iot and machine learning: towards industry 4.0 empowerment. IEEE Access 9, 115429–115441 (2021). https://doi.org/10.1109/ACCESS.2021.3105297
22. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. **39**(2), 273–315 (2008). https://doi.org/10.1111/j.1540-5915.2008.00192.x