

Production Planning and Scheduling Using Machine Learning and Data Science Processes

Paulo Henrique DE MODESTI^{a,1}, Ederson CARVALHAR FERNANDES^a and Milton BORSATO^a

^a*Universidade Tecnológica Federal do Paraná, Curitiba, Parana, Brazil*

Abstract. Increasing manufacturing efficiency has been a constant challenge since the First Industrial Revolution. What started as mechanization and turned into electricity-driven operations has experienced the power of digitalization. Currently, the manufacturing industry is experiencing an exponential increase in data availability, but it is essential to deal with the complexity and dynamics involved to improve manufacturing indicators. The aim of this study is to identify and allow an understanding of the unfilled gaps and the opportunities regarding production scheduling using machine learning and data science processes. In order to accomplish these goals, the current study was based on the Knowledge Development Process – Constructivist (ProKnow-C) methodology. Firstly, selecting 30 articles from 3608 published articles across five databases between 2015 and 2019 created a bibliographic portfolio. Secondly, a bibliometric analysis, which generated comparative charts of the journals' relevance regarding its impact factor, scientific recognition of the articles, publishing year, highlighted authors and keywords was carried out. Thirdly, the selected articles were read thoroughly through a systemic analysis in order to identify research problems, proposed solutions, and unfilled gaps. Then, research opportunities identified were: (i) Big data and associated analytics; (ii) Collaboration between different disciplines; (iii) Solution Customization; and (iv) Digital twin development.

Keywords. Machine Learning, Production Planning, Data Science, Predictive Analytics, Scheduling.

Introduction

Technological development in recent decades has provided a huge amount of data. This dataset is known as Big Data [1]. The manufacturing industry is experiencing several advantages from this data availability [2]. Some of these benefits turned out in higher quality and productivity [3]. On the other hand, it is essential to deal with the dynamics involved in it. Despite the benefits and opportunities, there are many risks to improve manufacturing indicators that cover qualified professionals and complexity [4].

¹ Corresponding Author. paulomodesti@alunos.utfpr.edu.br.

The development of computer science tools offers high potential to change the manufacturing domain. Some of these developments are applied to machine learning (ML), which is the ability of computers to solve problems without being programmed to do so [5]. Although computer engineering has been such a challenge for many manufacturing companies, the absence of its full potential application can increasingly affect their competitiveness level.

Strong candidates to surpass most of the major challenges of complex manufacturing systems are ML algorithms. This data-driven approach is capable of solving highly complex, non-linear patterns in data of several types and sources and transforming raw data into models, which are then applied to prediction, and in this case, production scheduling [5].

The application of ML techniques has increased over the last years due to several factors, such as the availability of complex data with poor transparency [6] and the increased convenience and power of ML tools. Many ML algorithms are developed to analyze large amounts of data and are capable of handling high dimensionality [7].

This paper presents a literature review for ML methods and data science processes applied to the production scheduling in manufacturing. The main contribution of this research is to identify which trends and proposed solutions have been developed from the last five years as well as to find unfilled gaps and opportunities of research in this field.

This research is organized into five sections. The first section is the introduction; the second section presents the methodological aspects; the third section describes the research development; the fourth section discusses the main conclusions and, finally, the fifth one exhibits the final considerations.

1. Methodological aspects

According to [8], a methodological framework consists of an examination and description of the approaches used in the research, in order to define the set of methods, procedures, and assumptions relevant to the subject. The lack of a standard that allows the adoption a single procedure to establish the research methodology allows the choice of the methodological framework to vary according to the researcher's knowledge [9].

In addition, [10] stated that a research can be classified according to both its purpose and design. Goal-based research is divided into three groups: exploratory, descriptive and explanatory. The design-based study contains two main groups: paper-source or people-source research. Paper-source research has two categories: bibliographic and documentary. People-source has several categories as experimental, ex-pos-facto, gathering, case study, action research, and participant. Considering the goal of this research, it can be labeled as exploratory, as it aims to provide greater familiarity with the topic, in order to make it more explicit or make hypotheses. It can be said that exploratory research has as main objective the improvement of ideas or the discovery of intuitions. In addition, as the source of information is based on material already prepared, composed of scientific articles, the research was classified as bibliographic. Then, the approach adopted proposed by [10] is summarized in Figure 1.

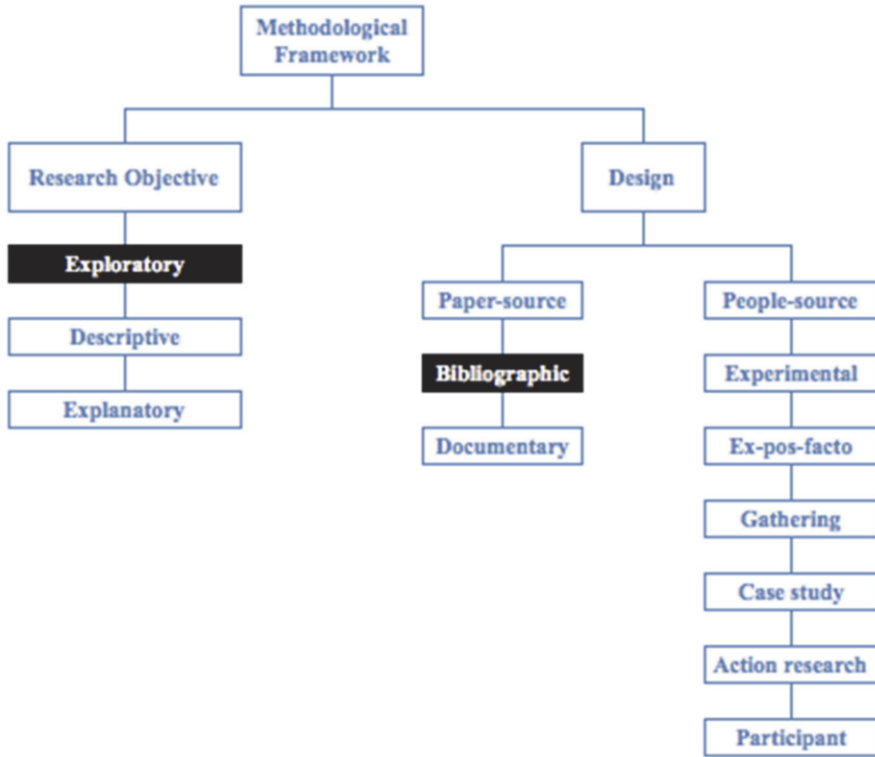


Figure 1. Methodological structure.

The development of the study was based on the Knowledge Development Process – Constructivist (ProKnow-C) methodology, proposed by [8]. This methodology is divided into four steps: (i) bibliographic portfolio selection; (ii) bibliometric analysis; (iii) systemic analysis and; (iv) definition of the research question and objective. The purpose of using this methodology is to facilitate the selection of relevant articles as it aims to identify references with high academic relevance in databases accessed through the Internet, using the bibliometric analysis to filter the selected articles.

The ProKnow-C methodology is adopted up to the systemic analysis step as the goal of this research is to identify trends and opportunities related to the topics under analysis. Access to research in databases was carried out through the CAPES journal portal. The software that supported the study was Mendeley Desktop © version 1.19.4 and Microsoft Excel for Mac 2011 version 14.3.9.

2. Research development

The research process begins with a problem that will motivate the search for data to be examined. The bibliographic portfolio along with bibliometric and systemic analysis, are the main steps towards achieving the research goals [10]. The portfolio includes all relevant published content for the main topic and analysis is vital to discover new

opportunities and gain a deep understanding of production scheduling using machine learning and data science processes.

2.1. Bibliographic portfolio

The bibliographic portfolio contains scientifically recognized publications from 2015 to 2019 and encompasses the most developed stage of knowledge on the subject so far.

In this study, three axes were chosen to be analyzed: Data Science, Production Planning and Machine Learning. In order to have a precise search, 18 keywords were defined to better match the axes analyzed.

The Boolean operators “AND” and “OR” were used to find articles associated with the keywords. The combination used was: ("data analytics" OR "data mining" OR "complex event processing" OR CEP OR "predictive analytics" OR "data acquisition") AND ("Production planning" OR scheduling OR "mixed-integer program*" OR MILP OR "monitoring system*" OR prediction OR "smart manufacturing" OR "intelligent manufacturing") AND ("regression learning" OR "supervised learning" OR "pattern classification" OR "supervised classification" OR "artificial intelligence" OR "machine learning").

There are several scientific databases for searching publications, but only five were selected due to a preliminary analysis of both their relevance in manufacturing research and the availability of full access. They were: Web of Science, Wiley, ProQuest, Scopus and IEEE (Figure 2).

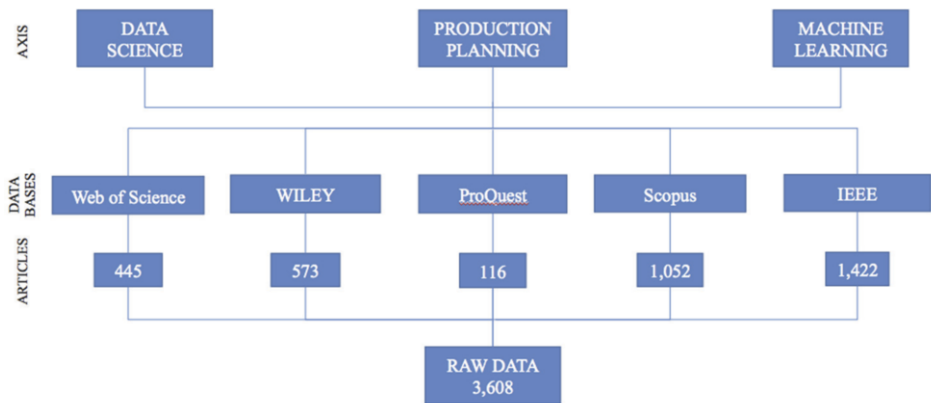


Figure 2. Research scheme.

The result obtained by primary research in the databases had a total of 3608 publications. For getting the best relevant articles from the bibliographic portfolio, four consecutive filters were performed for the current raw data: Published in related Vehicles, Duplicate Elimination, Title Reading and Abstract Reading.

Then, a search was filtered for publication in related vehicles (i.e. in each database to select articles exclusively and in specific areas according to the research goal), such as Production Operations Management, Operations Research, Computer Science, Engineering and Decision Sciences. Therefore, the outcome was 1241 raw articles.

Thus, Mendeley was used to help find the most relevant articles across the following three filters as Duplicate Elimination, Title Reading and Abstract Reading.

Mendeley is a free reference manager, which is used to facilitate access and analysis of each article. The main information of the portfolio such as authors, title, year of publication and journal were downloaded directly from the databases to Mendeley, in order to obtain a single set of data.

Firstly, a duplication check was performed in Mendeley, resulting in a reduction of 37 articles. Secondly, after reading 1204 article titles, the portfolio was reduced to 86 papers related to the topic. Thirdly, abstract readings of each remaining article were performed, resulting in 32 articles considered related to the research subject. The result of each filter is illustrated in Figure 3.

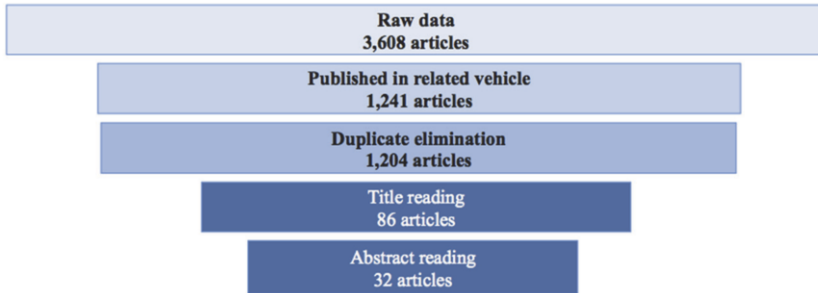


Figure 3. Filter actions and results.

2.2. Bibliometric analysis

Bibliometric analysis is related to the understanding and evaluation of articles in the bibliographic portfolio. This analysis was performed through charts containing portfolio information, such as impact factor, number of citations, publications by year, type of research and most of the published journals.

The portfolio impact factor, as shown in Figure 4, has a regular distribution, ranging from 1 to 6. Two journals can be highlighted with a higher impact factor.

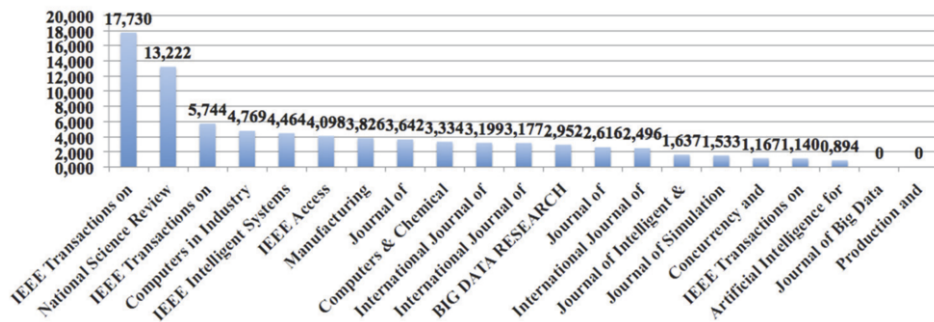


Figure 4. Impact factor of vehicle.

In Figure 5 is possible to see several articles with a large number of citations. Through this number of citations, it is possible to conclude that the portfolio has good relevance.

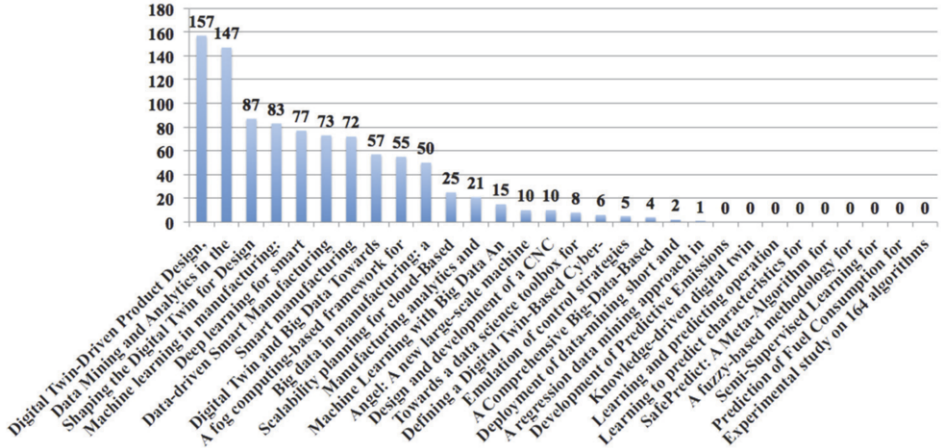


Figure 5. Number of citations distribution.

These two with more than 150 citations belonging to the authors Fei Tao and Zhiqiang Ge can be considered references in relation to the selected axes.

The increase in the number of articles per year in the last five years of research (Figure 6) shows the trends of the most discussed and published subjects. The slight reduction in the last year is explained by the fact this is the current year.

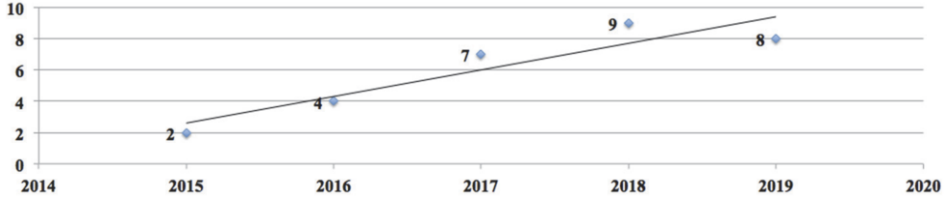


Figure 6. Number of articles published by year.

Analyzing the classification of the 30 types of articles, 77% of the articles are classified as proposed solutions, with the remaining percentage divided equally in evaluation research and philosophical paper (Figure 7).

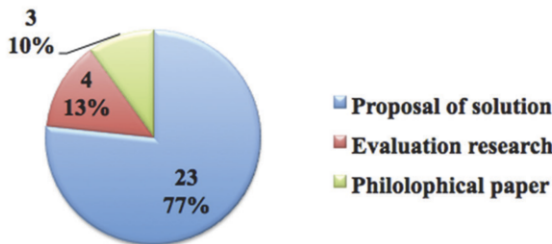


Figure 7. Type of research.

It is usually possible to regularly check the publication of an article by journal, but its differentiation has been identified in some journals. Figure 8 presents this statement. IEEE Access is the journal with the largest number of publications. In addition, most of the articles in this journal were printed in the current year.

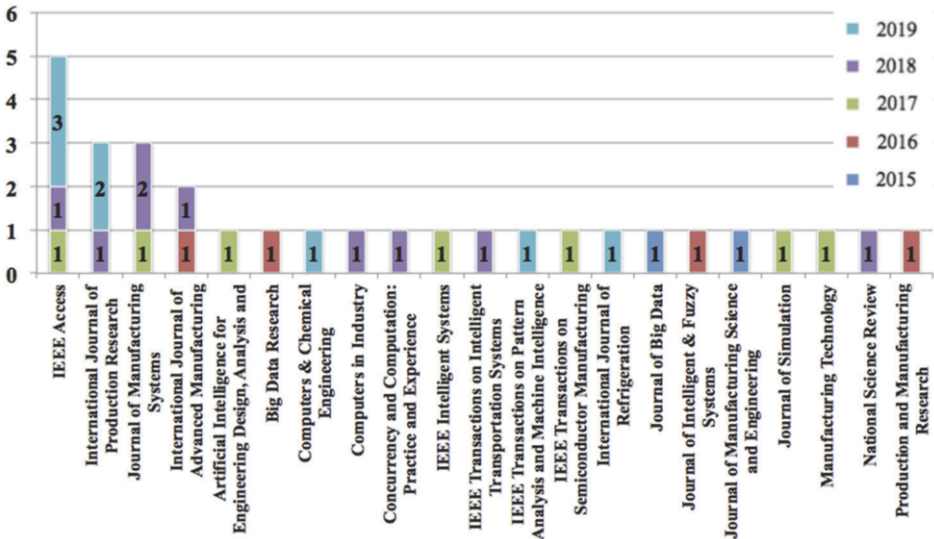


Figure 8. Number of publications by journal and year.

Conform to the results, 63% of the 30 published articles were classified as descriptive, and the others are opportunities for predictive and prescriptive developments (Figure 9).

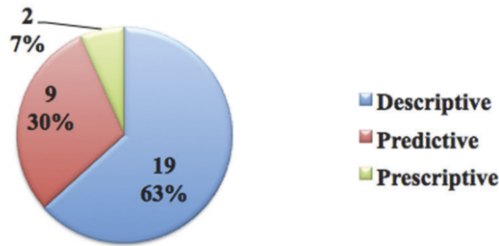


Figure 9. Classification of the research according to the result.

And last but not least, a comparative chart was created on published articles classified by the manufacturing area to analyze how manufacturing topics were divided between these studies. It was identified in four categories: Virtual Manufacturing, Process and Planning, Control, and Scheduling and Quality Management. Their results are visible in Figure 10.

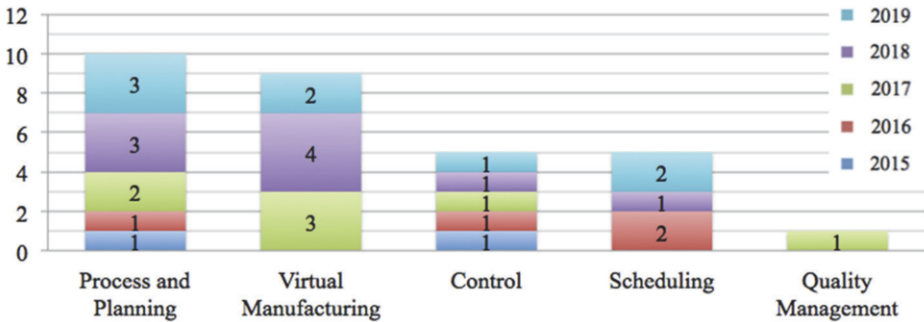


Figure 10. Published articles classified per manufacturing area and year.

2.3. Systemic analysis

The purpose of systemic analysis is to deeply study each article in the portfolio in order to find gaps in the literature. The criterion for finding gaps is based on the author's decision about what is relevant to the research [11]. The articles were read and analyzed in full in the following aspects: objective, context problem, methodology, results, trends, and opportunities.

And last but not least, after reading full-text filter, 12 articles were eliminated because they were not related to the research subject. The remaining articles resulted in 20 articles. Then, an extra review of the article references was carried out and another 17 articles were analyzed. From this review, 7 were eliminated and 10 added to the final portfolio. The final result of the number of articles in the portfolio is 30 articles.

2.3.1. Identified research problems

The main problems found by the authors were highlighted as follows.

Large datasets: several studies in the portfolio reveal that there are many algorithms capable of solving production planning problems [12]. On the other hand, most of these algorithms take a long time to solve problems or fail when they work with large datasets.

External factors: combining the use of different algorithms, it is possible to solve many complex artificial intelligence problems, such as maintenance prediction or digital twin. On the contrary, it is important to note that external factors are not considered, such as economic recession, weather or an accident [13].

Find the best algorithm: artificial intelligence is a broad concept, where there are many different strategies to be adopt. One such strategy is ML, which has several algorithms. Therefore, an important problem published by some portfolio authors is the difficulty in deciding which algorithm is more appropriate to use.

2.3.2. Proposed solutions

Although the models proposed in the portfolio can identify relationships between variables, they may not indicate a causal relationship. However, causality is essential in

identifying true root causes. Therefore, due to the complexity of the topic, researchers are challenged to investigate [5].

As stated by [14], process and planning are the most researched area of Big data in manufacturing. Therefore, manufacturers must be able to understand the exponential increase in demands for data production, even adopting the analytical techniques necessary to extract the meaning of this data.

[15] stated both the digital twin and big data play important roles in promoting smart manufacturing. The digital twin allows manufacturers to access real-time information and validate data between physical and virtual equipment. Combined with accurate analysis and predictive resources, this technique can provide the manufacturer with a huge improvement in his productivity.

It is proposed by [16] that ML can act with Big Data analytics. The role of Big Data is to retrieve the desired data from the database, and ML is to predict information. By integrating the two techniques, it is possible to work with large datasets in different formats and provide a suitable solution for manufacturing.

Finally [17], intelligent manufacturing means using advanced data analytics to complement physical science and improve productivity and decision making. Deep learning serves advanced analytics tools to process and analyze manufacturing data. The development of deep learning technologies, a subset of artificial neural networks with various layers, and their advantages over traditional ML is clear. Therefore, computational methods based on deep learning improve system performance in manufacturing.

[18] presented a paper to provide a computational methodology based on industrial alarm analysis and an event-log database. Based on the results of frequent sequence mining. A novel methodology and associated network-based visualization technique have been published to learn how the alarm messages that occur represent the states of technology and can be applied to predict future events. In the end, a neural network was explicitly designed to capture long-term dependency on sequences, and a sequence-to-sequence deep learning model was proposed.

2.3.3. Unfiled gaps identification

For the raised problems and their solutions previously presented, some unfilled gaps can be identified.

- *Big data and associated analytics*: these will be the key technologies to extract the required knowledge and provide intelligence in all engineering processes. As proposed by [14], it is logical that future work should focus on the development of systematic reviews and literature related to the manufacturing areas identified in this study, such as creating a systematic big data review in the manufacturing diagnosis.
- *Collaboration between different disciplines*: the need for interdisciplinary studies related to artificial intelligence presents a great opportunity, but also a significant risk, as collaboration between different disciplines, such as Computer Science, Industrial Engineering, Mathematics, and Electrical Engineering, is necessary to boost progress, according to [5] and [15].
- *Solution Customization*: [19] states that manufacturing is undergoing a technological revolution that, in turn, should revolutionize the industry. On the

other hand, users will require more personalization and electronics features for all products. This will close the cycle between design, manufacturing, marketing, sales, and aftermarket fields of the manufacturing industry.

- Digital twin development: the Digital twin enables manufacturers to manage two-way real-time mapping between a physical object and its digital representation, which opens the way for deep cyber-physical integration. [20] In combination with a digital twin, intelligent data-driven manufacturing will become more responsive, adaptable, and predictive.

3. Discussion chapter

Figure 11 shows the trends identified in the studies based on the selected portfolio. Process and planning show an increasing trend of study over the next years. Scheduling and control can be considered a stable field since the number of published studies is constant. It is interesting to note that Virtual Manufacturing shows a decline, but it is possible to infer that some related articles have not yet been published, justifying the decline. There is only one article related to Quality Management, so it is not possible to perform any analysis.

There are several machine learning technologies available to researchers. Based on the portfolio, it is possible to assume that Artificial Neural Networks are the most used algorithm in the research field, with k-means in second place.

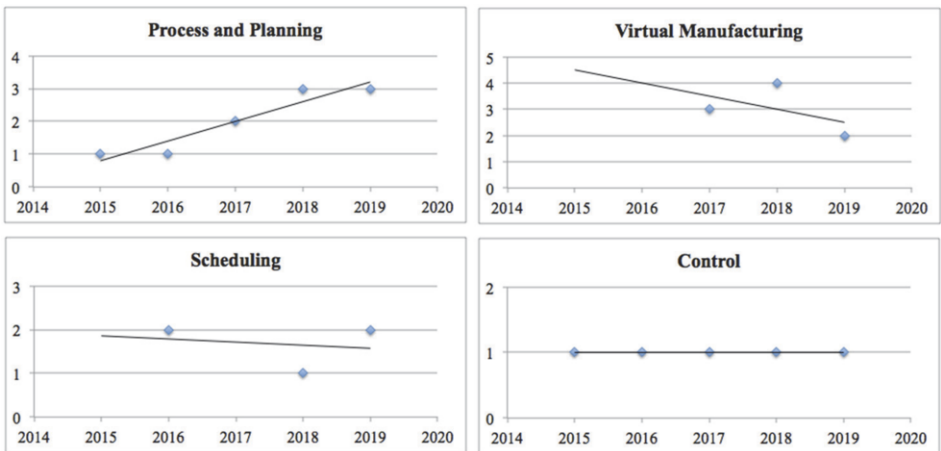


Figure 11. Published articles in manufacturing across the years.

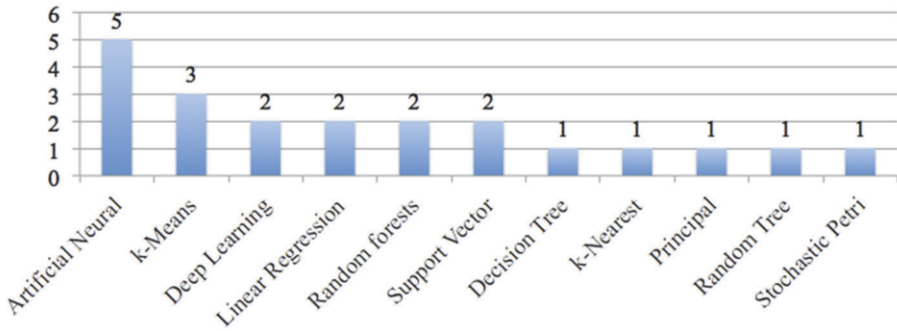


Figure 12. Types of machine learning more used in manufacturing.

4. Conclusions

The study does not attempt to exhaust the proposed subject, as there may be other sources to be investigated. Anyway, this work used a structured methodology designated ProKnow-C in a bibliographic and systemic analysis of production scheduling using machine learning and data science processes.

Firstly, the bibliographic portfolio was created, composed of 3608 published articles, found in five databases between 2015 and 2019. Then, 30 articles relating to the topic were selected.

Secondly, through a bibliometric analysis, the bibliographic portfolio was interpreted and evaluated. In addition, a comparison of the relevance of the journals in relation to the impact factor, scientific recognition of the articles, publishing year, highlighted authors and keywords was done.

Thirdly, in the systemic analysis the entire portfolio was thoroughly read, in order to identify the research problems, the proposed solutions, and the unfilled gaps. The research opportunities identified were: (i) Big data and associated analytics; (ii) Collaboration between different disciplines; (iii) Solution Customization; and (iv) Development of digital twins.

The use of the ProKnow-C methodology allows the researcher to have a deep understanding of the problems, goals, gaps, and opportunities under analysis. However, it depends on the researcher’s previous experience in selecting appropriate research axes and keywords in order to provide the best result.

In future studies, the intention is to perform the Big data analysis associated with ML, and focusing on scheduling of the production process, since the use of these tools tends to be widely used in the industry.

References

- [1] J. Lee, E. Lapira, B. Bagheri, and H. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manuf. Lett.*, vol. 1, pp. 38–41, 2013.
- [2] S. Chand and J. F. Davis, "What is smart manufacturing?," *Time Magazine*, 2010.
- [3] M. Elangovan, N. R. Sakthivel, S. Saravanamurugan, B. B. Nair, and V. Sugumaran, "Machine learning approach to the prediction of surface roughness using statistical features of vibration signal acquired in turning," *Procedia Comput. Sci.*, vol. 50, pp. 282–288, 2015.
- [4] T. Wuest, *Identifying product and process state drivers in manufacturing systems using supervised machine learning*. New York, NY: Springer Verlag, 2015.
- [5] T. Wuest, D. Weimer, C. Irgens, and K. D. Thoben, "Machine learning in manufacturing: Advantages, challenges, and applications," *Prod. Manuf. Res.*, vol. 4, no. 1, pp. 23–45, 2016.
- [6] A. Smola and S. V. N. Vishwanathan, *Introduction to machine learning*. Cambridge: Cambridge University Press, 2008.
- [7] K. Yang and J. Trew, *Multivariate statistical methods in quality management*. New York, NY: McGraw-Hill, 2004.
- [8] J. E. Tasca, L. Ensslin, S. R. Ensslin, and M. B. M. Alves, "An approach for selecting a theoretical framework for the evaluation of training programs," *J. Eur. Ind. Train.*, vol. 34, no. 7, pp. 631–655, 2010.
- [9] S.M. Petri, "Modelo para apoiar a avaliação das abordagens de gestão de desempenho e sugerir aperfeiçoamentos: sob a ótica construtivista" Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.
- [10] A. C. Gil, *Como Elaborar Projetos de Pesquisa*, 3rd ed. Belo Horizonte: Atlas, 1996.
- [11] L. ENSSLIN, S. R. ENSSLIN, and H. M. PINTO, "Processo de investigação e análise bibliométrica: avaliação da qualidade dos serviços bancários," *Revista de Administração Contemporânea*, pp. 325–349, 2013.
- [12] M. Usman and S. Asghar, "An Architecture for Integrated Online Analytical Mining," *J. Emerg. Technol. WEB Intell.*, vol. 3, no. 2, 2011.
- [13] R. Bragança, F. Portela, and M. Santos, "A regression data mining approach in Lean Production," *Concurr. Comput. Pract. Exp.*, p. e4449, Mar. 2018.
- [14] P. O'Donovan, K. Leahy, K. Bruton, and D. T. J. O'Sullivan, "Big data in manufacturing: a systematic mapping study," *J. Big Data*, vol. 2, no. 1, 2015.
- [15] Q. Qi and F. Tao, "Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018.
- [16] M. N. Rahman, A. Esmailpour, and J. Zhao, "Machine Learning with Big Data An Efficient Electricity Generation Forecasting System," *BIG DATA Res.*, vol. 5, pp. 9–15, 2016.
- [17] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.
- [18] G. Dorgo and J. Abonyi, "Learning and predicting operation strategies by sequence mining and deep learning," *Comput. Chem. Eng.*, vol. 128, pp. 174–187, Sep. 2019.
- [19] P. Lade, R. Ghosh, and S. Srinivasan, "Manufacturing analytics and industrial Internet of Things," *IEEE Intell. Syst.*, vol. 32, no. 3, pp. 74–79, 2017.
- [20] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *J. Manuf. Syst.*, vol. 48, pp. 157–169, 2018.