Research Article

Application of Big Data Analytics and Related Technology in Maintenance – Literature Based Research

Prof. (Dr.) T. Muthukumar

Professor (Business Analytics) & Associate Dean (Academic)
Xavier Institute of Management & Entrepreneurship (XIME) – Bangalore.
E-mail: *prokumar62@yahoo.com* ; Mobile: 7338941581

Abstract

Keywords

maintenance;
data analytics;
big data;
production systems;
manufacturing
systems

Digitalization is argued to increase the efficiency of maintenance activities in a production system. One consequence of digitalization is data deluge; this allows data analytics methods and technologies to be used. However, the actual data analytical methods and technologies used may differ, thus leading to many scientific papers on this topic. The purpose of my contribution is to find and cluster scientific papers regarding the implemented approaches relevant for use in production maintenance. My research is based on a broad, systematic literature review consisting of a two-step search approach combined with additional filtering and classification. Based on the search results, we evaluate and visualize the potential impact of data analytics on the subject of maintenance. The results of this study broadly summaries the research activities in production maintenance, whilst indicating that the impact of data analytics will grow further. Specific methodological approaches are clearly favored.

1. Introduction

Maintenance is increasingly important in modern manufacturing and production systems, as plant downtimes caused by machine or equipment damage may result in high repair costs owing to expensive technology and lost productivity [1–3]. Globalization adds to this issue, as unreliable deliveries due to problems in manufacturing processes may result in the loss of contracts and customers. Thus, a high degree of asset availability is a necessity in modern manufacturing companies [4]. Digitalization refers to a next-generation of maintenance paradigm aiming to use sensors, manufacturing execution systems (MES) and so on to collect data. This collection of digital data enables the use of statistical and machine-learning methods to improve productivity in the maintenance process [5].

Preventive maintenance still remains as a standard approach for a huge number of manufacturers [6]. This means that maintenance is carried out after a specific time or after a certain number of process steps have been completed by a piece of equipment or machine. However, this method does not allow for the actual condition of the changed parts and can result in relatively high costs, as components may be replaced while still in good working condition. On the other hand, having malfunctions due to components wearing out before their scheduled maintenance is a costly option [7].

A modern manufacturing company's processes generate vast amounts of data from its interconnected information systems, machines and equipment [8]. Moreover, the historical data stored in databases and real-time sensor data from production processes can be analyzed and used as a source of information and knowledge. This can improve many aspects of the company, including maintenance.

Data analysis can be carried out by applying different types of analytic methods and using a variety of different technologies, including data mining, exploratory analysis, machine learning. Depending on the way the data is analyzed and the knowledge intended to be gained from the analysis, the approaches can be classified as descriptive, diagnostic, predictive or prescriptive [9,10].

There is already a huge number of scientific papers concerning the application of data analytics and related technologies in maintenance [11–13]. However, there is no research effort reported to systematically review the current state of this new wave of industrial revolution. Therefore, the purpose of this paper is to find and cluster those relevant to maintenance use in a production context. This aim of the paper is to provide a systematic literature review, within data analytics covering the maintenance of production systems. More specifically, seven research questions are listed below. They provide a more appropriate answer as to the purpose of this paper:

1. How many publications have been published in different maintenance contexts and how does this number change over time?
2. Which analytical methods were used?
3. How has the use of analytical methods evolved? 4.Which technologies were used?
5. Which attributes of big data could be found?
6. What are the applications of data analytics in a maintenance context? 7.Which types of analytics are used in maintenance planning?

The findings of this study can be used as a basis for future research into the application of data analytics in production maintenance. Thus, to allow classification, the basics of data analytics, its methods and technologies must be defined. These are described in Section2. Once the basics had been defined, a suitable research and filtering approach had to be chosen. This is described in Section3, with the research questions of this paper. The results of the research are then visualized and explained in Section4. The answers to the research questions are derived from these results and shown in Section5. Finally, general conclusions are drawn regarding the application of big data analytics in maintenance.

2. Relevant Terms

This chapter introduces the terms "big data" and "data analytics" and their definitions are discussed in detail.

2.1. Big Data

To examine the application of big data analytics in maintenance, the term must be defined. As its name suggests, it is the application of analytical methods and techniques to big data. Big data can be defined by the "three Vs": volume, velocity and variety. These essential characteristics of data all present challenges for the methods and techniques of data analytics [14]. They can be defined as follows:

⌐ Volume describes the amount of data. This includes the size of a single record as well as the quantity of records [14].

⌡ Velocity includes the rate at which data is recorded and rate at which it must be processed [15].

⌡ Variety describes the differences in the data, especially in context of data structure [14].

2.2. Data Analytics

Data analytics is the application of mathematical methods to gain information from data, which can then be used to optimize processes or support decisions [15]. Figure1shows a reasonable classification of data analytics in a landscape of relevant keywords. Furthermore, the term "data analytics" has been further specified so that it can be used to classify the examined papers. This paper subdivides data analytics into data analytics methods and techniques. Thus, the method describes which general type of data analytics has been used (descriptive, for example). The term "technique" represents the specific approach used for the data analytics method. Different methods could therefore be defined as follows:

⌡ Descriptive analytics indicates applications of data analytics in describing and understanding situations based on past and present data [16].

⌡ Diagnostic analytics is an application of data analytics to investigate the causes and effects of situations [17].

⌡ Predictive analytics is an application of data analytics that use data and mathematical concepts to show the relationship between data, in order to predict future outcomes based on changes in the dataset [16].

⌡ Prescriptive analytics describes the application of data analytics using mathematical models to create a set of complex alternatives from the available data. This is then used to prescribe the best possible solution [16].
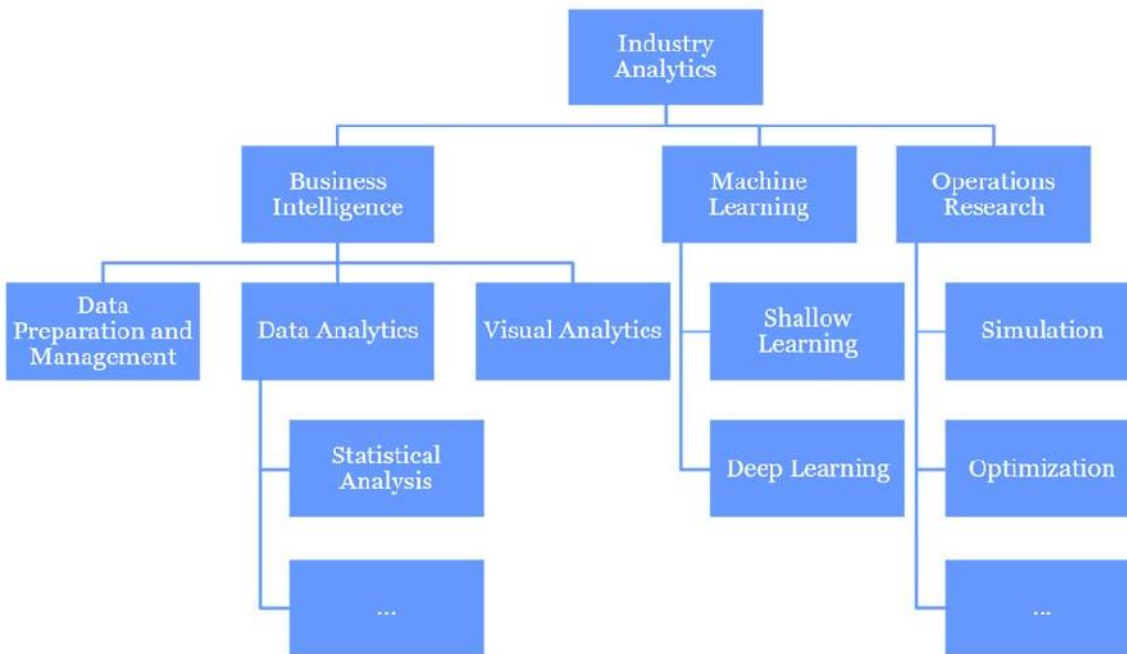

Figure 1. General classification of data analytics in a landscape of relevant keywords.

The following data analytics techniques may also be defined:

⌡ Machine learning: the design and study of algorithms which infer the function they compute from the sample data. In other words, machine learning can learn from the data and adapt to the changes and progress made, without the need for explicit programming [18].

⌡ Statistics: this unifies methods to condense, describe and evaluate data, thus helping to create a summary of (large) volumes of data. Thus, statistics provide more comprehensible information about data and support the drawing of conclusions [19].

⌡ Simulation: this means imitating complex real-world systems by constructing a mathematical model, which can then be evaluated numerically. Thus simulation affords the opportunity to estimate the behavior and characteristics of a system in certain scenarios [20].

⌡ Optimization: this process comprises independent variables and a measure of "goodness" (objective function), depending on the variables. At the end of an optimization, the combination of certain variable values leads to an "optimal" objective function value [21].

3. Principles and Methods

3.1. Search Strategy

This research is based on an organized literature review. To this end, a systematic approach must be chosen [22]. We conducted the study so as to search for the keywords in journal titles, keyword lists and abstracts. I have focused on papers where the authors classified their work in the area of big data, thus excluding works where this term was not used. As shown in Figure2, the search strategy for this review consists of two major steps and a subsequent filtering process:
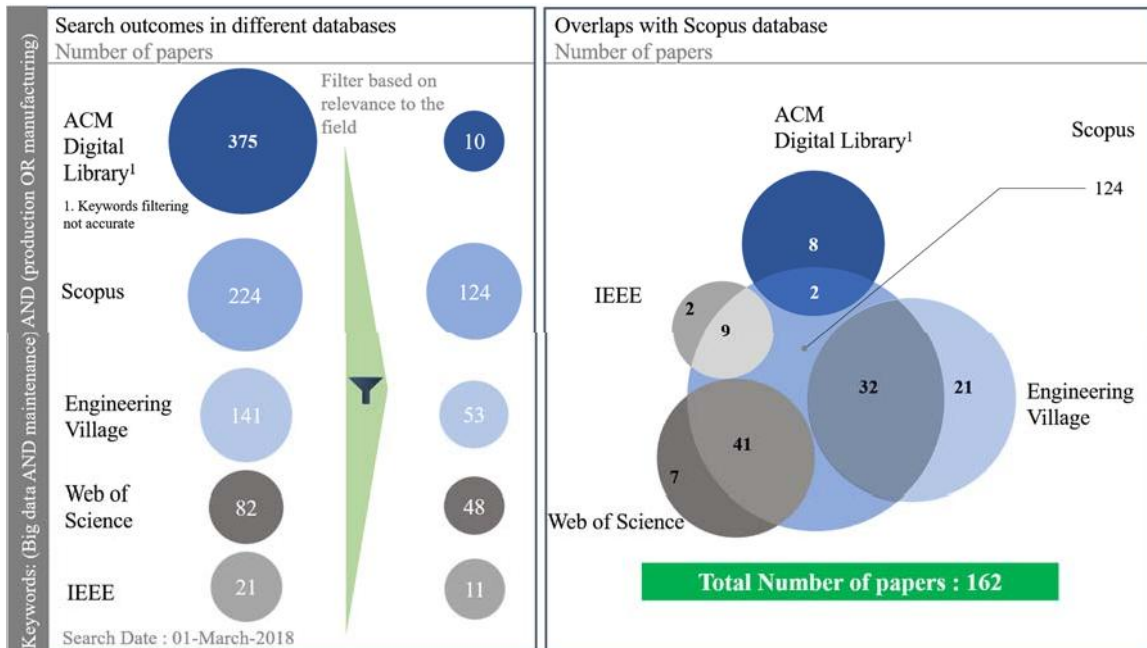

Figure 2. Search process and total number of papers.

Firstly, relevant keywords were chosen, based on the aim and scope of the research. These keywords were: big data, maintenance, production, manufacturing. The keywords were then constructed as a search string using the operator OR and AND in between them ((big data AND maintenance) AND (production OR manufacturing)). The reason for selecting the string "big data" is that it can be considered an overarching term allowing the use of data mining or other data science techniques to analyses the data. Thus, using "big data" captures as many papers as possible. Moreover, our evaluation of different keywords showed that the papers which appeared more relevant to this study had the term "big data" in their abstracts' keywords list.

I have also excluded the phrases "data science" and "data mining" as "big data" is seen as an overarching term that enables the usage of data mining or other methods under data science techniques to analyses the data. Thus, the rationale behind selecting the string "big data" is to capture as many pieces of research as possible.

Secondly, appropriate data sources were selected to collect scientific papers. For this study, I have chosen the ACM Digital Library (ACM—Association for Computing Machinery), Scopus, Engineering Village, Web of Science and IEEE Xplore (IEEE—Institute of Electrical and Electronics Engineers) databases, as they contained sources from a wide range of scientific disciplines. Only scientific papers published online before the end of February 2018 were collected.

3.2. Search Results, Filtering and Classification

Using the keyword combinations presented brought up 853 papers. These papers were then filtered by inclusion of the relevant subject areas that are related to maintenance with production/manufacturing. Those subject areas are: Business, Management and Accounting, Computer Science, Decision Science, Engineering, Mathematics and Material Science. The other areas such as medicine, biology, chemistry and so on are excluded, as they are not related to maintenance within manufacturing and production (see Figure 2—left). The remaining search results have been cleaned of duplicates, which lead to 162 possibly relevant publications (see Figure 2—right).

Based on the abstracts and keywords given in the scientific papers themselves, the results have been filtered further to ensure that only publications relevant to answering the research questions are considered. The 162 papers have been further filtered manually by reading their full texts and assessing their relevancy to the research questions. This resulted in 122 papers.

To be able to answer the research questions, the search results were classified using the classifiers listed in Table 1, which are based on the research questions. Not every paper could be classified according to all classifiers. Therefore, the concrete numbers of each figure in the 'Results' section might vary.

Table 1. Applied classifiers.

Classifier	Reason	Research Question
Title	Used as an identifier	-
Author	Used as an identifier	-
Year	Used to identify trends on the popularity of the subject over time	1
Taxonomy	Used to identify the used data analytics methods	2, 3, 6, 7
Technology	Used to identify the used data analytics technologies	4, 6
Big data attributes (Velocity, Variety, Volume)	Used to further elaborate on the possible attributes	5

4. Results

4.1. Publications Per Year

Figure 3 shows the growth in the number of relevant publications from the first paper identified in this study, from 1996 until today. While there were fewer than four publications from 1996 to 2012, it should be noted that there has been a steady increase in relevant publications starting in 2013 and peaking at 25 published papers in 2016. Since then, the number of papers published has decreased. It is noteworthy that the proportion of papers dealing with very specific product-related maintenance has risen. These papers often address specifics of maintenance that are related to the design of the product, rather than the actual maintenance or data analytics-

based maintenance of the corresponding manufacturing processes. These papers have therefore been filtered out, leading to a decrease in the overall number of papers. It should also be mentioned that some of the 2017 contributions may not have been fully indexed at the time the databases were searched (February 2018).

4.2. Applied Methods and Techniques

To get an overview of the data analytics methods and techniques that were applied, the relevant papers have been classified accordingly. It should be noted that publications could be considered for more than one type of analytics. Figure 3 shows the selected papers in a timeline of their publication date.
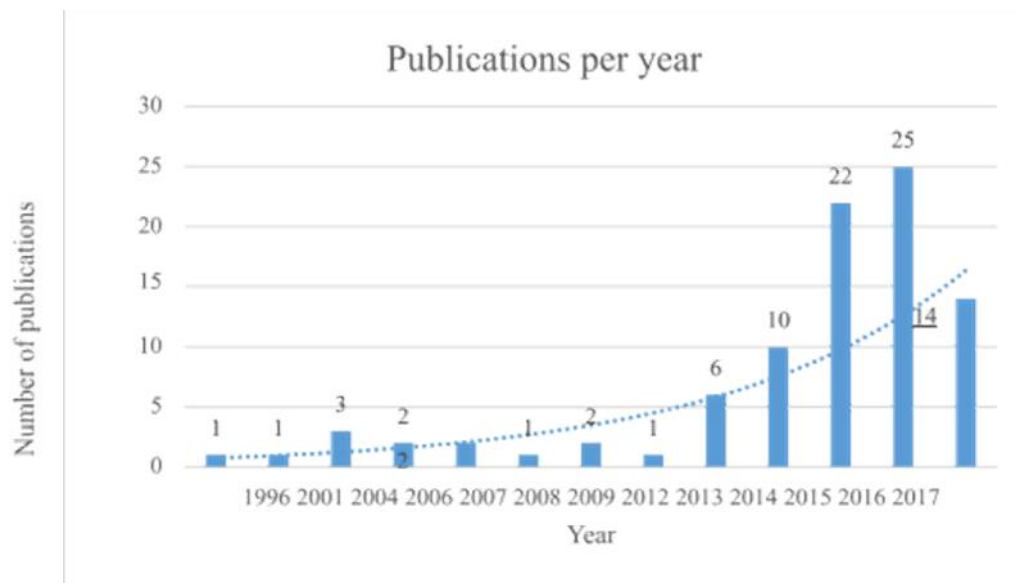

Figure 3. Publication numbers for data analytics in maintenance, per year.

Figure 4 shows the number of publications belonging to the different data analytics methods. It should be noted that four papers could be assigned to two different methods and one paper incorporated three methods. These papers have been added to each method.
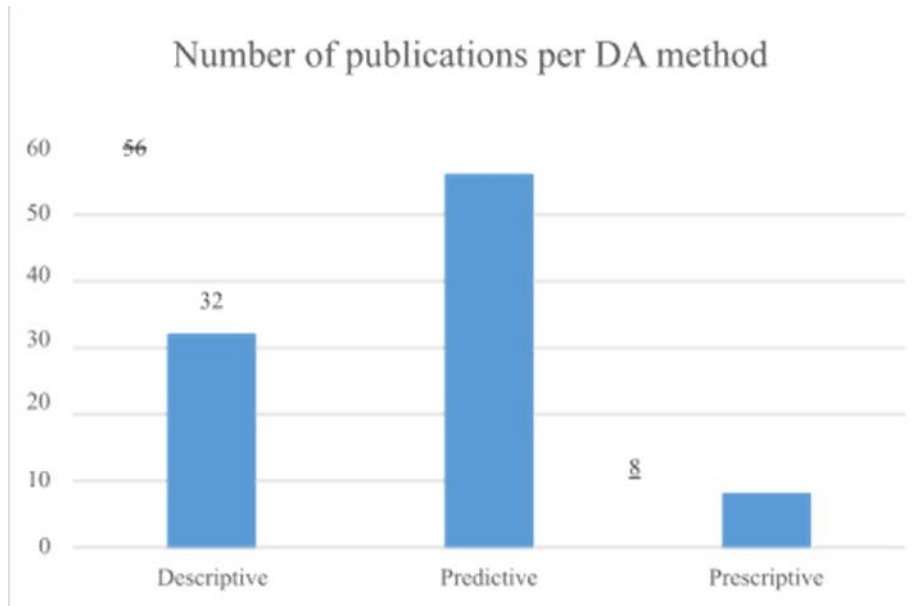
Figure 4. Number of publications per data analytics method.

Figure 5 shows the number of publications per year that involve the different data analytics methods. Consistent with Figure4, if different methods are included, then papers are added to each method category.
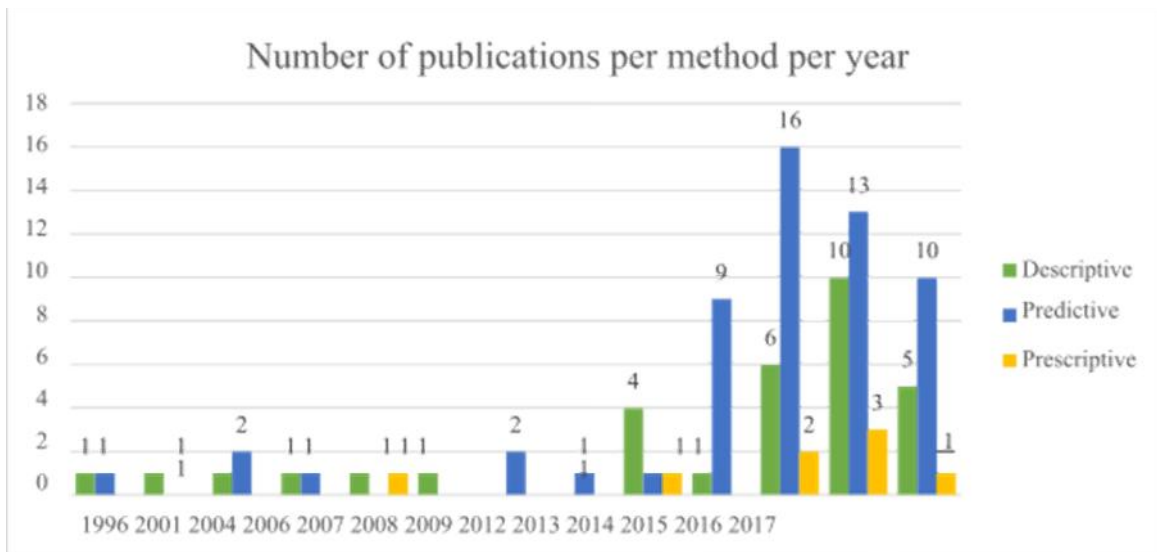

Figure 5. Publication numbers for data analytics of maintenance, per year.

Figure 6 illustrates the number of publications which refer to one or more techniques. A number of papers could not be simply classified to a single technique and were thus added to each technology used. Also, 14 publications could not be classified under one of the introduced techniques. In these cases, the authors had not given a clear description of the techniques used; in other cases, they used specific analytical approaches that could not be matched.
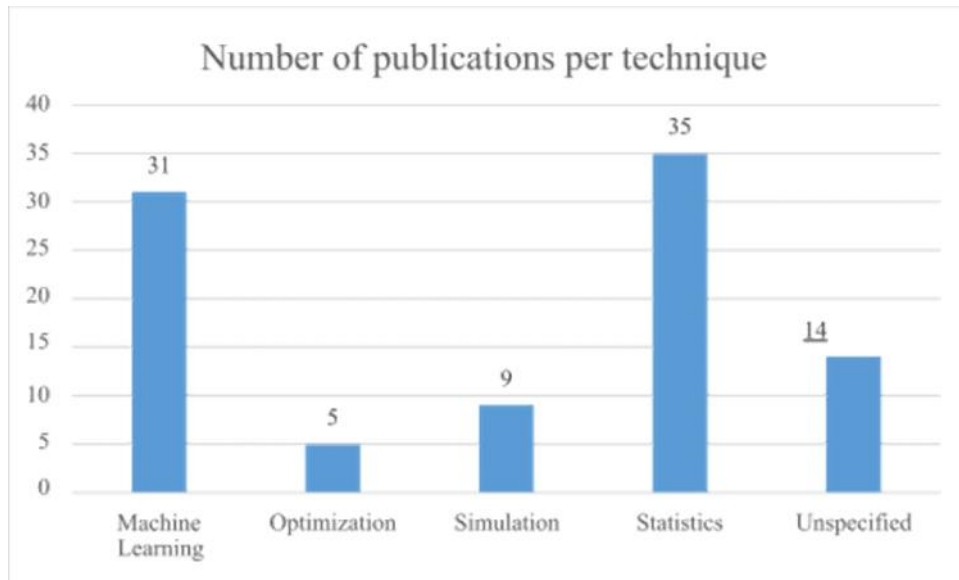
Figure 6. Techniques referenced in the relevant publications.

Figure7 gives an overview of the different data analytics techniques applied to the different methods. As before, papers that could be classified under multiple data analytics techniques were added to all of these.
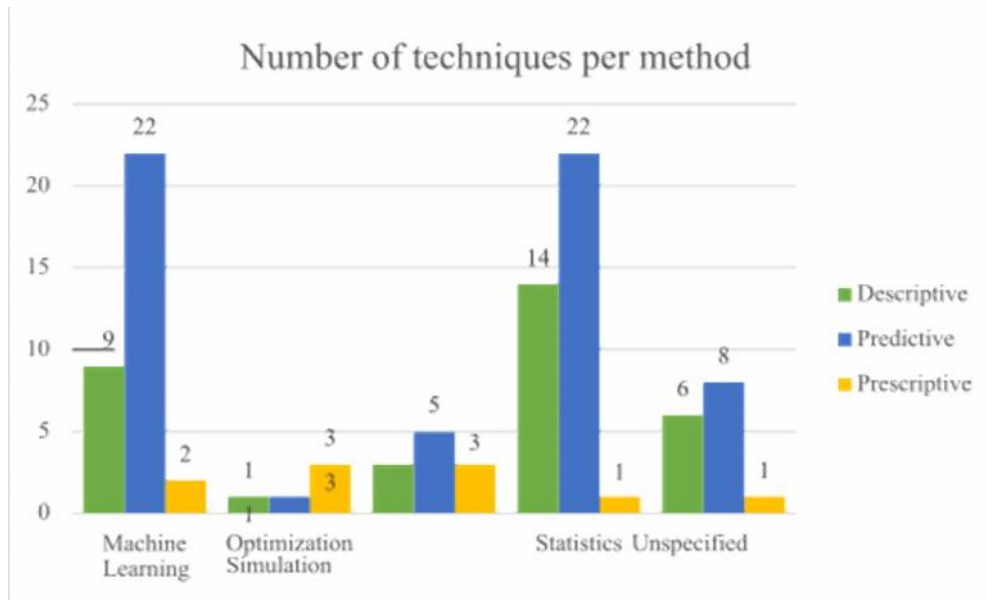

Figure 7. Data analytics techniques used within the different data analytics methods.

5. Discussion of the Research Questions

Question 1. How many publications have been published in different maintenance contexts and how does this number change over time?

The first relevant publication dated back to 1996. Even though the term "big data" was introduced early 2000, the paper of 1996 can be classified under big data from today's view regarding the technology at that time and the approach made. The number of publications was relatively low, with a maximum of three per year until 2012. The number of publications per year increased from 2013 onwards and peaked at 25 in 2016. It should be noted that only papers up to December 2017 were considered (indexed up to February 2018). It is likely that this number will change during the year. Overall, it can be seen that research efforts in big data analytics in the field of maintenance within production or manufacturing is growing, year on year.

Question 2. Which analytical methods were used?

Three of the stated analytical methods (descriptive, predictive and prescriptive) were used. Predictive analytics took the largest share with 63 publications, followed by 34 publications dealing with descriptive analytics and 10 for prescriptive analytics. Four papers could not be clearly identified as "descriptive or predictive" or "predictive or prescriptive". Diagnostic analytics was not mentioned explicitly in the papers evaluated. Thus, it can be assumed that it was used in some approaches, at least as a basis for further analytics and reuse within the predictive models. As this is just a presumption, the papers concerned were not added to diagnostic analytics. Overall, there was a lack of prescriptive analytics, as evidenced in the results. However, this may have been due to the complexity of designing the prescriptive analytical frameworks, as compared to the descriptive and predictive ones, because it involves a combination of optimization approaches and decision-making theories. Developing prescriptive analytics algorithms for

maintenance applications is a future research direction.

Question 3. How has the use of analytical methods evolved?

Descriptive and predictive data analytics were both mentioned in the first relevant publication, whereas sharp increases could be observed in 2013 (descriptive) and 2014 (predictive). Prescriptive analytics was first addressed in a relevant publication in 2007. The number of publications for prescriptive analytics rose slightly in 2015 and 2016. It can be observed that the increase of predictive analytics was the highest, followed by descriptive analytics. The predictive analytics increase can be attributed to development of predictive algorithms in other fields (such as machine learning) and its application to real-world problems. Moreover, the fact that the figure for descriptive analytics was less may be due to its inherent inclusion when designing the predictive algorithms. The increase in prescriptive analytics was the lowest in terms of the total number of publications. It is comparable to the increase in descriptive and predictive analytics within a ten-year range of these terms being first addressed.

Question 4. Which analytical techniques were used?

The idea behind this question was to identify the extent to which analytics is being used in maintenance within production/manufacturing. The results show that most papers employed some form of analytics. Statistics was generally the most-used technique, with 36 publications mentioning it. Machine learning followed closely behind, with a total of 32 publications using this technique. Simulation and optimization were further behind with a total of 11 and six mentions respectively. A total of 14 publications could not be assigned to any specific analytical technique.

Taking the combination of analytical methods and techniques into consideration, this distribution is transferable to descriptive and predictive

analytics. Machine learning was used nine times in descriptive analytics, whereas in predictive analytics machine learning was used as often as statistics. This may indicate that the interdisciplinary research community (in terms of collaboration between computer scientists and maintenance) is maturing and evolving when it comes to finding application areas for machine-learning techniques. Simulation was applied four times for descriptive analytics and seven times for predictive analytics. Optimization was the least-used technique, with one publication in descriptive analytics and two in predictive analytics. Within prescriptive analytics, the distribution of data analytics techniques differed. The leading techniques used were optimization and simulation, each mentioned in three publications. These were followed by machine learning (two publications) and statistics (one publication).

Moreover, additional research is probably carried out by other authors, especially in the field of simulation and optimization for manufacturing maintenance without mentioning the listed keywords "data analytics" or "big data". This might arise from different research communities, where simulation experts see their work as applicable to operations research only. Considering this, even more applications in the predictive and prescriptive method category are reasonable.

Question 5. Which attributes of big data could be found?

The three Vs of big data could be found in most papers, starting with the huge volume of data used as the basis for analytics. We initially intended to classify the existing literature on an ordinal scale for each of the three Vs. One problem that arose during extraction of the papers was that very few authors give concrete numbers on the amount of data and data sources they had considered when writing their papers.

Volume aspect: the rationale behind analyzing the volume of data was to quantify how much was being used to develop analytical models. In other words, how much data was required for the models to work and how much training data for all types of learning algorithms were available. Only a few authors give details in their contributions, some citing the volume of historical data. For example, one contribution stated that they used the previous 60 days of data to build a predictive model.

Velocity aspect: the idea behind mapping research efforts to this attribute of big data was to get an idea of how the analytical models cope with the speed of incoming data (data aggregation is a common technique used to handle the velocity aspect, among other things). We hoped that best-practice solutions could be identified in order to give clear hints to other research teams. As with the volume aspect, most authors gave very few details on this aspect of their work. This aspect needs to be looked further by reviewers of those contributions, since validation and verification of results is made impossible.

Variety aspect: it can be seen from the analysis that most of the research efforts were focused on analyzing the data extracted from one type of information system (manufacturing execution system (MES) analytics, computerized maintenance management system (CMMS) analytics, logistics information system and so on). Though this could be seen from a variety aspect of big data taxonomy, there is a lack of research that combines all these types of data to make an effective prediction. Thus, one future direction might be to integrate data from different sources to develop better predictive models. As we know from our own practical experience, mapping different data sources, even from within a single company, often leads to huge problems. Unique identifiers are seldom used across all the information systems in a location.

Question 6. What are the applications of data analytics in a maintenance context?

Based on the results, it can be said that analytics were mainly applied to examine historical maintenance data and understand the efficiency and impact of maintenance work practices. A

significant number of papers used analytics to summaries the historical data on maintenance. 11 papers referred to predictive maintenance in their title and it was clear that they lean more towards maintenance planning. The most striking aspect of this result was that analytics were focused on enabling data-driven decisions. Some papers reported results from real-world industries, which shows that these industries recognize the importance of data-driven decisions.

Question 7. Which types of analytics are used in maintenance planning?

The publications in question showed that all the stated data analytics methods were used in maintenance planning, often in combination. Descriptive analytics was mainly used as a basis for predictive analytics by detecting prior maintenance interval trends and using this information to predict trends accordingly. Prescriptive analytics was then used to optimize maintenance intervals towards cost efficiency or reduced downtimes. Though most of the efforts were aimed at some kind of data-driven decision support, no contributions could be identified which addressed a fully automated solution in the sense of "autonomous maintenance". Since humans are needed for the maintenance task itself, the final decision on the "next action" was obviously intended to remain a manual one.

6. Conclusions

The aim of this research was a literature-based evaluation of the application of big data analytics in maintenance. Therefore, seven research questions were formulated, in context of which the literature-based research was conducted. A systematic approach was chosen for the research process, including a search and filtering strategy. The resulting papers were categorized based on defined classifiers and then visualized. Based on these categorized papers, it was possible to answer the research questions defined at the beginning.

In summary, there is an increasing trend towards applying big data analytics to maintenance. It is noticeable that different data analytics methods and technologies were used in varying combinations, even though predictive analytics appears the most popular in recent years. Statistics and machine learning are the most commonly applied when it concerns technologies. Alongside a general trend towards applying big data analytics to maintenance, a specialization trend can be observed.

In considering the systematic literature analysis presented in this paper, the following limitations are to be noted. Firstly, the search strings consisted of the words "big data", "maintenance", "production" and "manufacturing" and did not consider other related terms, such as "factory", "industry 4.0" and so on. Thus, the results reported in the paper should be interpreted in respect of these keywords. Secondly, the papers were collected from the largest databases such as Scopus, Web of Science and IEEE Xplore. There may be other papers which were not indexed in these databases but are still relevant to the research questions addressed in this paper. These papers have not been considered. Finally, only papers written in English (including the title, abstract and full text) were considered in this study; others were excluded. Therefore, to answer the research questions more comprehensively, papers from other databases and other languages should be included. Despite these limitations, this paper reports the current state-of-the-art of big data analytics within maintenance and suggests potential research directions.

The reference list of papers considered for analysis in this paper is available on request from the corresponding author.

References

1. Ahuja, I.P.S.; Khamba, J.S. An evaluation of TPM implementation initiatives in an Indian manufacturing enterprise. *J. Qual. Maint. Eng.* **2007**, *13*, 338–352. [CrossRef]

2. Cooke, F.L. Plant maintenance strategy: Evidence from four British manufacturing firms. *J. Qual. Maint. Eng.* **2003**, *9*, 239–249. [CrossRef]

3. Fraser, K.; Hvolby, H.-H.; Tseng, B. Maintenance management models: A study of the published literature to identify empirical evidence. A greater practical focus is needed. *Int. J. Qual. Reliab. Manag.* **2015**, *32*, 635–664. [CrossRef]

4. Cooke, F.L. Implementing TPM in plant maintenance: Some organizational barriers'. *Int. J. Qual. Reliab. Manag.* **2000**, *17*, 1003–1016. [CrossRef]

5. Bokrantz, J.; Skoogh, A.; Berlin, C.; Stahre, J. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. *Int. J. Prod. Econ.* **2017**, *191*, 154–169. [CrossRef]

6. Fraser, K.; Hvolby, H.-H.; Watanabe, C. A review of the three most popular maintenance systems: How well is the energy sector represented? *Int. J. Glob. Energy Issues* **2011**, *35*, 287–309. [CrossRef]

7. Ahuja, I.P.S.; Khamba, J.S. An evaluation of TPM initiatives in Indian industry for enhanced manufacturing performance. *Int. J. Qual. Reliab. Manag.* **2008**, *25*, 147–172. [CrossRef]

8. Subramaniyan, M.; Skoogh, A.; Salomonsson, H.; Bangalore, P.; Gopalakrishnan, M.; Sheikh Muhammad, A. Data-driven algorithm for throughput bottleneck analysis of production systems. *Prod. Manuf. Res.* **2018**, *6*, 225–246. [CrossRef]

9. Shao, G.; Shin, S.J.; Jain, S. Data analytics using simulation for smart manufacturing. In Proceedings of the Winter Simulation Conference, Savanah, GA, USA, 7–10 December 2014; pp. 2192–2203.

10. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.; Wuest, T.; Weimer, D.; Thoben, K. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]

11. Lee, J.; Kao, H.-A.; Ardakani, H.D.; Siegel, D. *Intelligent Factory Agents with Predictive Analytics for Asset Management*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 341–360.

12. Lee, J.; Ardakani, H.D.; Yang, S.; Bagheri, B. Industrial Big Data Analytics and Cyber-physical Systems for Future Maintenance & Service Innovation. *Procedia CIRP* **2015**, *38*, 3–7. [CrossRef]

13. Kejela, G.; Esteves, R.M.; Rong, C. Predictive analytics of sensor data using distributed machine learning techniques. In Proceedings of the International Conference on Cloud Computing Technology and Science, Singapore, 15–18 December 2014; pp. 626–631. [CrossRef]

14. Russom, P. Big Data Analytics. In *TDWI Best Practices Report 2011, Fourth Quarter*; The Data Warehousing Institute: Renton, WA, USA, 2011; pp. 6–9.

15. Runkler, T.A. *Data Analytics: Models and Algorithms for Intelligent Data Analysis*, 2nd ed.; Springer Vieweg: Wiesbaden, Germany, 2016; p. 1, ISBN 9783658140755.

16. Delen, D.; Demirkan, H. Data, information and analytics as services. *Decis. Support Syst.* **2013**, *55*, 359–363. [CrossRef]

17. Fleckenstein, M.; Fellows, L. *Modern Data Strategy*, 1st ed.; Springer: Cham, Switzerland, 2018; p. 133. ISBN 9783319689937.

18. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. *Machine Learning—An Artificial Intelligence Approach (Volume I)*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 1983; pp. 5–6, ISBN 9780934613095.

19. Taylor, J.K.; Cihon, C. *Statistical Techniques for Data Analysis*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; pp. 19–22, ISBN 9781584883852.

20. Law, A.M. *Simulation Modeling and Analysis*, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2014; pp. 1–6. ISBN 9780073401324.

21. Gill, P.E.; Murray, W.; Wright, M.H. *Practical Optimization*, 1st ed.; Academic Press Limited: San Diego, CA, USA, 1981; pp. 1–3. ISBN 9780122839504.

22. David, B.; Brereton, P. Performing systematic literature reviews in software engineering. In Proceedings of the 28th International Conference on Software Engineering, Shanghai, China, 20–28 May 2006.