systems

MDPI

Article

Data Science Supporting Lean Production: Evidence from Manufacturing Companies

Rossella Pozzi *, Violetta Giada Cannas and Tommaso Rossi

School of Industrial Engineering, Università Carlo Cattaneo—LIUC, 21053 Castellanza, Italy; vcannas@liuc.it (V.G.C.); trossi@liuc.it (T.R.)
* Correspondence: rpozzi@liuc.it

Abstract: Research in lean production has recently focused on linking lean production to Industry 4.0 by discussing the positive relationship between them. In the context of Industry 4.0, data science plays a fundamental role, and operations management research is dedicating particular attention to this field. However, the literature on the empirical implementation of data science to lean production is still under-investigated and details are lacking in most of the reported contributions. In this study, multiple case studies were conducted involving the Italian manufacturing sector to collect evidence of the application of data science to support lean production and to understand it. The results provide empirical proof of the link and examples of a variety of data science techniques and tools that can be combined to support lean production practices. The findings offer insights into the applications of the traditional lean plan–do–check–act cycle, supporting feedback on performance metrics, total productive maintenance, total quality management, statistical process control, root cause analysis for problem-solving, visual management, and Kaizen.

Keywords: lean production; data science; data analytics; Industry 4.0; PDCA

check for updates

Citation: Pozzi, R.; Cannas, V.G.; Rossi, T. Data Science Supporting Lean Production: Evidence from Manufacturing Companies. *Systems* **2024**, *12*, 100. https://doi.org/10.3390/systems12030100

Academic Editor: Marco Ardolino

Received: 7 February 2024
Revised: 10 March 2024
Accepted: 13 March 2024
Published: 15 March 2024

1. Introduction

Lean production (LP) is a set of management principles and practices geared toward eliminating waste in the manufacturing process, increasing the flow of activities that add value to the product from the customer's perspective, and helping companies achieve high efficiency and effectiveness to strive toward operational excellence [1]. Despite its advantages, the implementation of LP practices presents challenges and limitations that include the necessity to manually collect and analyse data and the need to check the project's status [2].

Recent studies focused on linking LP to digital technologies commonly under the label of 'Industry 4.0' (I40), and the positive link between these technologies and LP has been conceptually discussed by several researchers [3,4]. In the I40 context, data play a fundamental role [5,6], and operations management research is dedicated to exploring data science (DS)-related technologies such as big data analytics [7]. The integration of LP practices with DS techniques and tools is meant to enhance decision support systems in process improvement, thanks to the sharing and analysis of real-time data from several devices [8,9], which facilitates the monitoring and control of processes [10]. The availability of real-time data and the ability to use it through advanced analytics make it possible to improve productivity, responsiveness, and decision-making [11]. LP practices, such as JIT, Kanban, value stream mapping, Kaizen, and total productive maintenance, are reported to benefit from data analytics thanks to continuous flow monitoring, greater transparency of processes, fact-based decision-making, and planning [9].

Even though the link between DS and LP enables manufacturing companies to exploit the value of data to facilitate the application of traditional LP practices and boost their performance [12,13], studies addressing the integration of DS and LP are mainly theoretical, and most reported contributions lack details on how to implement them [14–16]. On the other hand, clarifying how I40 technologies affect LP practices would support practitioners in addressing this link [3,17] and would be relevant to motivate technology implementation in LP practices [15,18–20].

This study therefore explores the link between DS techniques and tools and LP practices and aims to fill the identified gaps by answering two research questions (RQs):

RQ1: Which DS techniques and tools support LP practices in manufacturing companies?
RQ2: How do DS techniques and tools support LP practices in manufacturing companies?

The first RQ's objective is to identify the DS techniques and tools that support LP practices in manufacturing companies, thus providing empirical proof of this link. The second RQ aims to define a model to guide the application of DS techniques and tools to LP practices. To achieve this goal, the present work is grounded on an exploratory, qualitative research approach [21] that involves a cross-case comparison of six cases.

The remainder of this paper is organised as follows. Section 2 presents the background of the study. Section 3 presents the case study research strategy, company sampling criteria, and data collection procedures. Section 4 presents the cross-case analysis, and Section 5 discusses the results. Finally, Section 6 highlights the contributions and limitations of the study and the directions for future research.

2. Background

DS techniques and tools range from Internet of Things (IoT) applications exploiting sensors to machine learning clustering and classification. According to [16], these techniques and tools can be classified based on the activities that they contribute to, such as (i) data gathering, (ii) data preparation, (iii) data representation and transformation, (iv) data exploration, (v) data computing, (vi) data modelling and analytics, and (vii) data visualisation and representation. These seven DS activities represent a set of sequential fundamental steps required to successfully transform massive amounts of raw data into meaningful knowledge. Data are considered an enabler for improvement and the ability to use data through advanced analytics improves operational performance, reducing waste in manufacturing [15].

Considering LP as the main strategy to reduce waste in manufacturing [1], LP practices are the techniques and tools for LP implementation, ranging from the widely adopted workplace housekeeping and standardised work to supplier involvement in design. In the literature, the set of LP practices has been organised in categories, named 'bundles'. Recently, the work by Bai et al. [22] proposed a complete list of practices classified into five bundles: three considering internal LP practices (i.e., production planning and control, process technology, and workforce) and two covering external LP practices (i.e., supplier and customer). Regarding the implementation of LP practices, the plan–do–check–act (PDCA) cycle is reported as the framework for process improvement, problem-solving, and LP implementation [23,24]. The 'Plan' stage considers the modelling and sets the objectives of the practice implementation. The 'Do' stage regards the management of the practice. In the 'Check' stage, the implementation compliance with the Plan stage is evaluated. Last, the 'Act' stage is devoted to improving the implementation.

Considering the improvements offered by DS techniques and tools applied to LP practices, a few studies proposed or presented practical applications of DS to specific LP practices, mainly considering a narrow set of DS techniques and tools and were limited to the 'production planning and control' LP bundle [16]. Among them, the study by Pozzi et al. [16] identified specific DS techniques and tools in each DS activity that can support the LP practice in the different stages of the PDCA cycle. For example, DS tools for data gathering, such as Radio-Frequency Identification (RFId), Pad, and indoor Global Positioning System (GPS) and DS techniques for data modelling and analytics, such as machine learning, deep learning algorithms, and image recognition, are proposed to 'support translating elementary motions into activities and setting manual task time standard for setup reduction' (see page 15, [16]) in the setup reduction practice. On the other hand, this study does propose a practical evaluation of the link. Also, the narrow range of studies addressing the improvement in LP practices by applying DS techniques and tools are mainly theoretical and lack practical cases [14–16].

Hence, this work aims to empirically investigate and validate this link, by adopting a multiple case study methodology, which is recognised as being suitable for exploratory investigations, providing insights that could inform propositions for further theory development and more effective practice [21,25,26]. Figure 1 reports a list of DS techniques and tools for each DS activity based on [16], the list of LP practices organised in the bundles based on [22], and their link through the PDCA cycle, i.e., the three concepts this work refers to.

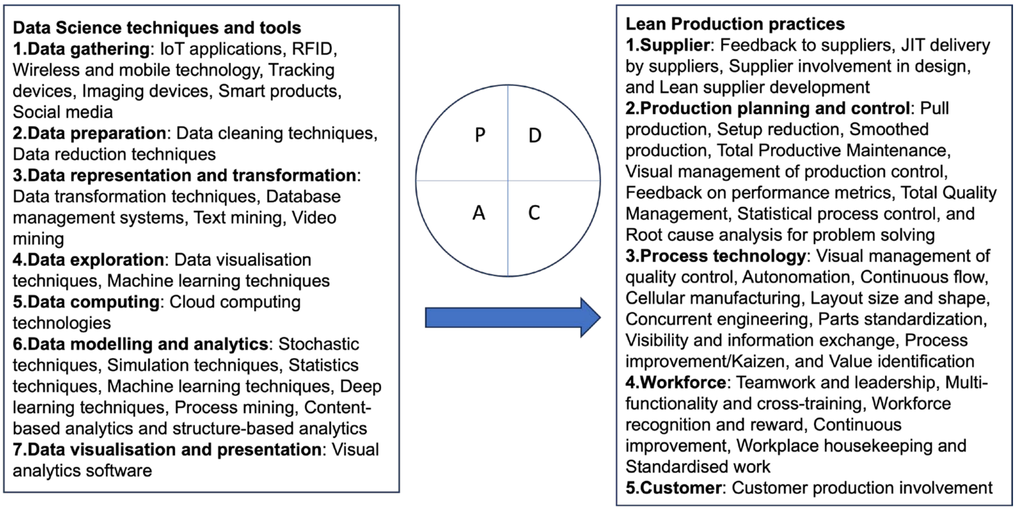

Figure 1. DS techniques and tools and LP practices, linked in the stages of the PDCA cycle.

3. Materials and Methods

Considering the focus of this study, the sample of interest consists of Italian manufacturing companies that have adopted one or more DS techniques and tools to support LP practices. In all cases, the adoption of LP practices occurred before the advent of I40, and DS adoption was aimed at supporting LP. The companies selected represent heterogeneous cases in terms of firm size and manufacturing industries.

Table 1 provides data on the sampled companies, using code names to protect their identities. The column "Industry" shows the companies' classification according to their ATECO codes [27]. ATECO codes are a system of economic activity classification adopted by the Italian National Institute of Statistics (ISTAT) to identify the main economic activity of a company. The "Number of employees" and "Turnover M€" columns show information derived from the AIDA database [28], a database of information on Italian companies such as their history, financial performance, and ownership structure. The column "Size" shows the companies' size, which is defined based on the definition of small, medium and large enterprises provided by the European Commission [29], which defines small enterprises as ones that have a turnover lower than or equal to 10 M€ and less than 50 employees, and medium enterprises as ones that have a turnover lower than or equal to 50 M€ and less than 250 employees. Finally, the last two columns present the "Number of informants" that were interviewed in each company and their "Job position", by their contractual job descriptions. The interviewees selected for the cases were key informants with a lead position in the organisation or the implementation process under analysis and were initiators/responsible for the DS implementation process.

Table 1. Case study overview.

Case	Industry	Number of Employees	Turnover (M€)	Size	Number of Informants	Job Position
A	Medicines and other pharmaceutical preparations	250	200	Large	2	• Industrial organizational method and portfolio manager • IT manager
B	Rubber component	270	80	Large	3	• Lean manager • Engineering and digitalization manager • Process systems engineer
C	Electrical equipment	36	10	Small	2	• Quality manager • Industry 4.0 intern
D	Electric motors	40	10	Small	2	• Quality engineer • Maintenance and Industry 4.0 engineer
E	Sun lenses	180	30	Medium	1	• Manufacturing excellence leader
F	Electrical equipment	240	40	Medium	2	• Lean process engineer • IT manager

As suggested by the literature, we collected data from three concurrent sources—interviews, documents, and direct observation—to triangulate the evidence [21,30]. Each of the three sources addresses different goals. Interviews can provide insightful explanations with a focus on the study topic, but they may also be limited in terms of factual details. Documents can provide facts but are subject to retrievability concerns [31]. Direct observation can support researchers when addressing discrepancies between what people say in interviews and casual conversations and their actions [32]. The interview questions were based on a semi-structured approach focused on (1) general information about the company and the interviewee, (2) LP practices applied by the company, and (3) DS techniques and tools employed by the company to support LP practices. The detailed case study protocol is reported in Table 2.

For each selected case, the authors conducted two rounds of interviews. In the first round, most of the data were collected through face-to-face semi-structured interviews and Gemba walks in the manufacturing facility, which allowed the researchers to visually evaluate DS adoption and an understanding of these implementations. In the second round, most questions were case-specific because they were intended to clarify DS adoption and its effects.

All the interviews were recorded. For each case, the authors generated a database consisting of interview transcripts, field notes, and archival data. To minimise interviewee bias, data collected from the interviews were compared with those from observations and documents. Working in two-person project teams that met regularly helped balance impartiality and involvement and limited the tendency to overidentify with interpretations. As suggested by [33], the data were analysed following a two-step procedure of within-case and cross-case analyses. In the within-case analysis, we developed one table for each case for qualitative data analysis, and, to avoid alternative explanations of the same outcomes and to maintain the chain of evidence, coding was performed by classifying the results within the theoretical framework. Then, in the cross-case analysis, we combined the tables to obtain the results of how DS was adopted in LP.

Table 2. Case study protocol.

Source 1: Face-to-Face Interview	
General information	Companies' approximate turnover, employees, industrial sector, competitive environment, interviewee/s role
Lean production practices	Regarding each lean production bundle (supplier, production planning and control, process technology, workforce, customer), what are the practices adopted by the company?
Data science tools and techniques	Regarding the seven data science activities (data gathering, data preparation, data representation and transformation, data exploration, data computing, data modelling and analytics, data visualisation, and presentation), what are the tools and techniques applied by the company?
Data science and lean production	• Which data science techniques and tools are applied to support lean production practices? • How do data science techniques and tools support lean production practices?
Source 2: Direct observations	
Plant tour	Direct observation of the production department during work shifts with the possibility of observing manufacturing and/or assembly activities and asking the employees and/or managers additional questions related to the processes, lean production practices applied, and data science techniques and tools implemented.
Source 3a: Official documents	
Company's website	Company information (history, strategy, mission, success factors, and others) and product information (product types, product features, technical data, applications, and others).
News and press	Up-to-date information related to recent business initiatives, new product launches, and new technology introductions.
National database	10 years of history-related information on Italian companies (balance sheet, number of employees, sector, and others).
Source 3b: Internal documents	
Documents (digital or paper)	Procedures, budgets, product catalogues, etc.
Tools	Data science tools are applied for data gathering, preparation, representation and transformation, exploration, computing, modelling and analytics, visualisation, and presentation.

4. Results

The results from the multiple case study research are presented in Table 3, aligned with the theoretical constructs this investigation draws upon, namely lean production (LP) bundles and practices, data science (DS) activities, techniques, and tools, and the stage of the plan–do–check–act (PDCA) cycle where the DS technique and tool were deployed to bolster the LP practice. As delineated in Table 3, the scrutinised cases predominantly leveraged DS in LP practices within the production planning and control (PPC) bundle, such as total productive maintenance (TPM), visual management of production control, feedback on performance metrics, total quality management (TQM), statistical process control (SPC), and root cause analysis for problem solving, in addition to the process technology bundle, which includes the visual management of quality control and process improvement/Kaizen. Subsequently, the results are elaborated by examining the role of DS in augmenting each LP practice considered in the case studies.

Table 3. Case study results.

Case	LP Bundle	LP Practice	DS Activity, Techniques, and Tools DS Activity Legend: (1) Data Gathering, (2) Data Preparation, (3) Data Representation and Transformation, (4) Data Exploration, (5) Data Computing, (6) Data Modelling and Analytics, (7) Data Visualization and Presentation	PDCA Stage
A	PPC	Feedback on performance metrics	(1): IoT applications; (3): Data transformation; Database management systems; Text mining; (4): Data visualisation techniques; (6): Statistics and Machine learning techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		TQM	(1): IoT applications; (6): Machine learning techniques; (7): Visual analytics software	ACT Improvement in quality management
		SPC	(1): IoT applications; (3): Database management systems; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
	Process technology	Process improvement/ Kaizen	(1): Wireless and mobile technology; (3): Text mining; (6): Statistics techniques; (7): Visual analytics software	CHECK Process monitoring
B	PPC	TPM	(1): Wireless and mobile technology; (6): Statistics techniques; (7): Visual analytics software	CHECK Monitoring autonomous maintenance progress
		Feedback on performance metrics	(1): IoT applications; Wireless and mobile technology; (3): Database management systems; Text mining; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		TQM	(1): IoT applications; (2): Data cleaning techniques; (3): Data transformation techniques; Database management systems; (4): Data visualisation techniques; (6): Process mining; (7): Visual analytics software	PLAN Modelling production process analytics
			(1): IoT applications; (2): Data cleaning techniques; (3): Data transformation techniques; Database management systems; (4): Data visualisation techniques; (6): Simulation tools; Statistics analytics; (7): Visual analytics software	DO Identification of variability
			(1): IoT applications; (2): Data cleaning techniques; (3): Database management systems; (5): Cloud computing technologies; (6): Machine learning techniques; (7): Visual analytics software	ACT Improvement in quality management
C	PPC	TPM	(1): IoT applications; (2): Data cleaning techniques; (4): Machine learning techniques; (6): Machine learning techniques; (7): Visual analytics software	PLAN Modelling production process analytics
			(1): IoT applications; (2): Data cleaning techniques; (3) Cloud computing technologies; (4): Machine learning techniques; (6): Machine learning techniques; (7): Visual analytics software	DO Fault detection
		Feedback on performance metrics	(1): Wireless and mobile technology (3): Database management systems; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		Root cause analysis for problem-solving	(1): Wireless and mobile technology: (6): Machine learning techniques	DO Extracting the root cause

Table 3. *Cont.*

Case	LP Bundle	LP Practice	DS Activity, Techniques, and Tools DS Activity Legend: (1) Data Gathering, (2) Data Preparation, (3) Data Representation and Transformation, (4) Data Exploration, (5) Data Computing, (6) Data Modelling and Analytics, (7) Data Visualization and Presentation	PDCA Stage
D	PPC	Visual management of production control	(1): IoT applications; (2): Data cleaning technique; (3): Database management systems (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		Feedback on performance metrics	(1): IoT applications; (2): Data cleaning techniques; (3): Database management systems; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		SPC	(1): IoT applications; (3): Data transformation techniques; Database management systems; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
E	PPC	Visual management of production control	(1): IoT applications; (2): Data cleaning technique; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		Feedback on performance metrics	(1): IoT applications; Wireless and mobile technology; (2): Data cleaning technique; (3): Database management systems; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		TQM	(1): Wireless and mobile technology; (3): Database management systems; (6): Statistics techniques; (7): Visual analytics software	DO Identification of variability
	Process technology	Visual management of quality control	(1): IoT applications; (2): Data cleaning technique; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
F	PPC	Visual management of production control	(1): IoT applications; Wireless and mobile technology; (2): Data cleaning technique; (3): Database management systems; Text mining; (5): Cloud computing technologies; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
		Feedback on performance metrics	(1): IoT applications; Wireless and mobile technology; (2): Data cleaning technique; (3): Database management systems; Text mining; (5): Cloud computing technologies; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field
	Process technology	Visual management of quality control	(1): IoT applications; Wireless and mobile technology; (2): Data cleaning technique; (3): Database management systems; Text mining; (5): Cloud computing technologies; (6): Statistics techniques; (7): Visual analytics software	ACT Providing up-to-date and real-time information in the field

4.1. DS Techniques and Tools Supporting TPM

Companies B and C support autonomous and planned maintenance activities using DS in different stages of the PDCA cycle. Company B applies DS to support the monitoring, i.e., the check stage, and boost autonomous maintenance activities. To this end, “after completing a maintenance activity, operators record the related data in Microsoft Forms modules via industrial pads. The collected data are then visualised through customised dashboards and reports generated using Microsoft Power BI software, representing the status of autonomous maintenance activities in a 2D plant layout". In Company C, DS supports the development of predictive maintenance activities for modelling the process analytics deriving the process models from event data, i.e., the plan stage, and fault detection, i.e., the do stage. Real-time machine status data are collected from the machines' programmable logic controllers (PLCs), whereas 'the sounds emitted by the machines are collected in real-time by microphones installed in the machine area'. These 'data are prepared using cleaning algorithms to remove outliers, missing, and wrong values'. Apache software is used 'to store the data in practical formats and to merge different databases'. In the plan stage, the process model is derived from recorded data adopting neural networks. In the do stage, decision tree and clustering are adopted 'to explore the stored data to understand what machine learning model can be applied to perform analytics on the link between machine noise, as a proxy of either its downtime or speed issues and following machine performance'. Dashboards support the visualisation of the process information.

4.2. DS Techniques and Tools Supporting Visual Management of Production Control

Companies D, E, and F adopt DS to provide the workforce with up-to-date and real-time information about production schedules and progress, i.e., the act stage. Manufacturing data are collected in real-time through IoT applications from machine PLCs that communicate via OPC-UA protocols, sensors, and forms that receive manual input. After collection, data are checked and cleaned using cleaning algorithms during the data preparation phase because 'data are often transmitted between two different systems, and this leads to errors' (Company F). Companies D and E prepare the data through a specific software implemented in the Manufacturing Execution System (MES) that 'makes it possible to create and use tools for data pre-processing and algorithms for data cleaning'. Companies D and F store data in an SQL database and ensure that they are 'ready to be managed and analysed' (Company D). Company F also applies text-mining techniques 'to extract keywords from text manually inserted by operators'. Company F uses cloud computing to perform computing activities at the machine level. All the companies perform statistical analyses. For example, Company D adopts Qlik View software to model and 'analyse data for the calculation of performance and quality indexes'. The data visualisation and presentation activity consists of enabling a 'customised dashboard to share the performance data with the "Andon" displays placed on the production lines' (Company D), 'the displays installed on production lines to monitor production and quality performance' (Company E), and 'dashboards and monitors to share within the production plant and, in particular, on production lines, real-time indicators about performance and quality' (Company F).

4.3. DS Techniques and Tools Supporting Feedback on Performance Metrics

All interviewed companies aim to achieve the real-time assessment of production metrics through key performance indicators (KPIs), such as machines' overall equipment effectiveness (OEE) or the lead time of production processes (Company B), i.e., the act stage of the implementation. In this context, the main sources of manufacturing data are IoT applications: machines' PLCs and IoT sensors. Data from machines' PLCs are transmitted using the OPC/OPC-UA protocol (Company A), edge computing systems (Companies B, D, E, and F), or wireless LAN controllers (Company C). IoT sensors installed in the manufacturing system are also 'exploited to obtain additional information on the production status' (Companies D and F). Wireless and mobile technologies are exploited by companies B, E, and F. Data preparation is undertaken by Companies D, E, and F in the same way as it is performed to support the visual management of production control, that is, through cleaning algorithms and MES software. Cleaning algorithms are adopted 'to check and correct data entered manually by operators, for example, when indicating the cause of machine downtime' (Company F). During the data representation and transformation phase, an SQL database is used 'to create an extensive and unique database merging data from machines and other sources’ (Company B). Company E adopts SQL databases to manage data. Company A performs the data representation and transformation using transformation, SQL databases, and text mining. Company C extracts data in pipelines with Apache Airflow and applies an Extract/Transform/Load process to place data from two sources in a MySQL database. Companies B and F also apply text-mining techniques and tools, such as Power Automate, ‘to extract keywords from text manually inserted by operators’ (Company B). The data exploration activity is reported only by Company A, which applies visualisation techniques such as histograms, scatter plots, historical time series, and radar charts. Cloud computing technologies are implemented by Company F ‘to support computing activities at [the] machine location’. All cases report data modelling and analytics, mainly through statistical analyses and tools such as Statgraphics software (Company A), Qlik View software (Company D), and Power BI (Company E). Moreover, Company A applies machine learning techniques to identify relationships and summarise the data. The data visualisation and presentation phase is supported by dashboards created using a variety of tools, such as Qlik Sense software (Companies A and D) and custom dashboards and cockpits. Company C adopts Linux dashboards to show real-time OEE.

4.4. DS Techniques and Tools Supporting TQM

Companies A and B support existing TQM implementations using DS in the act stage, i.e., improving quality management. Company A analyses data gathered from the machines through predictive modelling algorithms for timely defect identification. Company B gathers data in real-time from PLCs and accelerometers installed in the machines. Then, through R language, it applies ‘pre-processing data tools, robust algorithms, and statistical analysis to clean the data before starting the next step’, adopts ontologies for data representation and transformation, storing the data in practical formats ready for analysis, and conducts data exploration using decision trees, association rules, and regression functions ‘to detect causality relationships between variables and obtain a general interpretation of the data collected’. Collected data are modelled and analysed by using machine learning techniques to predict the output of production and to detect quality issues before they occur. Both companies engage in data visualisation and presentation using dashboards through Power BI and R Studio.

Additionally, Company B and Company E exploit DS to support new TQM applications. Company B, thanks to the benefits obtained in the act stage, decided to include DS for the plan and do stages of TQM applications, i.e., modelling production process analytics and identifying variability. It uses data collected from the machines, cleaned, reduced, transformed, and explored by using R language and R studio, to perform process mining, employing BPMN language, and to visualise results employing Signavio software. This allows Company B to produce process models, identify criticalities, and define where to address TQM actions and variability analyses, which are performed through statistics analytics, conducted using R language, to analyse historical data, identify feasible and objective statistical limits, and predict process outputs for process control. Moreover, advanced simulation tools (e.g., transport phenomena, discrete-event simulation with simmer tool, and uncertainty assessment) are used to enable quantitative decision-making for TQM investments. Company E gathers data related to product quality control from manual forms filled by the operators and read and recorded with mobile technologies. Then, it prepares data by applying data cleaning techniques with Power Query and applies descriptive statistics techniques, employing Power BI, ‘to extract meaningful information from the data collected related to the variability of the production output’.

4.5. DS Techniques and Tools Supporting Statistical Process Control

Companies A and D support the existing statistical process control practice by providing up-to-date and real-time information in the field, i.e., the act stage. The companies collect data from machine event logs and Extensible Markup Language (XML) files from PLCs. Both SQL (Company D) and NoSQL (Company A) databases are exploited to perform data representation and transformation. Company D uses the R programming language to analyse XML files and extract the necessary information next to the workstation without storing it in the company's central system. Company A uses Statgraphics software to perform statistical pattern analysis, such as process capability and analyses of variances, and to perform timely process temperature controls and achieve the desired energy exchange in the process. Company D uses statistical analyses, control charts, and machine learning models in R to suggest to operators when to correct line parameters. Company D conducts data visualisation and presentation by displaying the control chart on a monitor at the production lines. For this activity, Company A uses Qlik Sense software.

4.6. DS Techniques and Tools Supporting Root Cause Analysis for Problem-Solving

Company C supports the extraction of root cause analysis, i.e., the do stage, for the identification of machine cycle time variability causes. Cycle times are collected in real-time by exploiting machines' wireless LAN controllers. Clustering models are used to conduct data modelling and analytics, such as decision trees and clustering techniques, to investigate the factors that could influence cycle times.

4.7. DS Techniques and Tools Supporting Visual Management of Quality Control

Companies E and F apply DS to provide up-to-date and real-time information in the field of visual management of quality control, i.e., the act stage. Data are collected in real-time from machines (Company E and F) and forms where information is manually input (Company F). Data are cleaned, and text mining techniques are applied to the manual input. Data are then stored in SQL databases. In company F, cloud computing allows the calculation of statistical analyses at the machine level. The visualisation of quality metrics occurs through 'displays installed on production lines to monitor production and quality performance' (Company E) and 'dashboards and monitors to share within the production plant and, in particular on production lines, real-time indicators about performance and quality' (Company F).

4.8. DS Techniques and Tools Supporting Process Improvement/Kaizen

Company A uses DS to support the validation of process improvement/Kaizen practice, i.e., the check stage. Data on improvement ideas are provided by everyone in the organisation and are collected using Microsoft Forms. Power Automate software performs text mining on the proposed ideas and extracts information. Idea status, important dates, and application area are summarised through statistics techniques and are shown using Power BI analysis and the dashboard.

5. Discussion

This section discusses the results of the implementation of DS techniques and tools to support LP practices in manufacturing companies. It further compares these results with those of previous studies to answer the following two RQs.

RQ1. Which DS techniques and tools support LP practices in manufacturing companies?

The findings of this study agree with those of previous research that considers the use of DS to support LP practices, providing empirical examples in the PPC bundle [34], e.g., [35]. Moreover, as discussed by [16], among those reported by the interviewed companies, no empirical implementation was related to the supplier, customer, and workforce bundles of LP practices. Considering both the PPC and process technology bundles, not all practices emerge as supported by DS. By contrast, the results of this study present examples of empirical implementations not found in previous studies, such as the Kaizen and visual management of quality control practices of the process technology bundle. Considering the PPC bundle, the study reports the use of DS to support LP practices that have not been previously encompassed, such as autonomous maintenance in TPM and TQM.

Analysing in detail the DS implementations to support LP practices, some analogies and differences emerge from cases. The number of implementations found seems to not depend on the company size. Companies C and D, classified as "small", have reported the same amount of implementation and variety of techniques and tools as other bigger companies. On the other hand, the LP practices that the studied companies support using DS techniques and tools seem to depend on the company industry. Companies A, B, and E support TQM using DS. These companies, manufacturing medicines and other pharmaceutical preparations (Company A), rubber components for medical devices (Company B), and sun lenses for eyewear (Company E), all belong to the broader healthcare sector that is well known for requiring high-quality standards and procedures. Similarly, Company D, manufacturing electrical components for the automotive industry, supports SPC. Again, in this case, the assurance of quality is central in the field where the company competes and supplies components. Considering Company C and Company F, which manufacture electrical components for household appliances, their applications focus on time delivery by supporting the root cause analysis of variations in cycle times and machine availability with TPM (Company C), and the visual management of the shopfloor (Company F).

The activity that benefits most from DS in companies that have already adopted a set of LP practices, such as feedback on performance metrics, visual management of production control, visual management of quality control, and SPC, is the provision of up-to-date and real-time information in the field. Hence, a variety of techniques and tools are provided by the cases. For example, machines' PLCs and IoT sensors, the OPC/OPC-UA protocol, edge computing systems, and wireless LAN controllers are reported as alternatives to conduct the data gathering activity, providing examples to managers interested in supporting this activity in LP practices.

This study extends the previous literature and finds that a combination of DS techniques and tools needs to be implemented to achieve the desired results; however, not all DS activities need to be implemented. Across the seven DS activities, three consistently emerged from the cases: data gathering, data modelling, analytics, and data visualisation and presentation. Data gathering is typically performed through IoT applications and wireless and mobile technology, such as sensors, wireless LAN controllers, or industrial pads. Statistics techniques, machine learning techniques, process mining, and simulation tools are exploited to extract useful knowledge on patterns, relationships, models, or trends from data that constitute the core of the application. A variety of visual analytics software, such as Power BI, R studio, Linux, and Signavio, help to report results. Several data sources, including IoT applications, avoid the need to check data quality, remove anomalies and artifacts, and pre-process dirty, incomplete, or inconsistent data, that is, to undertake data preparation. Data computing is meant to develop a powerful distributed infrastructure to handle the analysis of extensive data that may be not needed in all cases, such as when central analyses of the data can support the analysis with no loss in effectiveness. Additionally, the search for useful features, basic properties, and the relationship that may exist between the variables through an exploratory data analysis performed by the data exploration activity can be avoided if the data analytics applied are already defined. This finding extends that of previous studies on empirical implementations.

RQ2. How do DS techniques and tools support LP practices in manufacturing companies?

This study confirms that the application of DS techniques and tools to support LP practices can be carried out within the traditional lean PDCA cycle, proposed by [16]. This finding agrees with the previous literature reporting applications in the do stage [34], e.g., [35], while extending it to present DS applications supporting the plan, check, and act stages. Figure 2 depicts the DS activities, techniques, and tools adopted by the analysed companies to support LP practices in the different stages of the PDCA cycle.

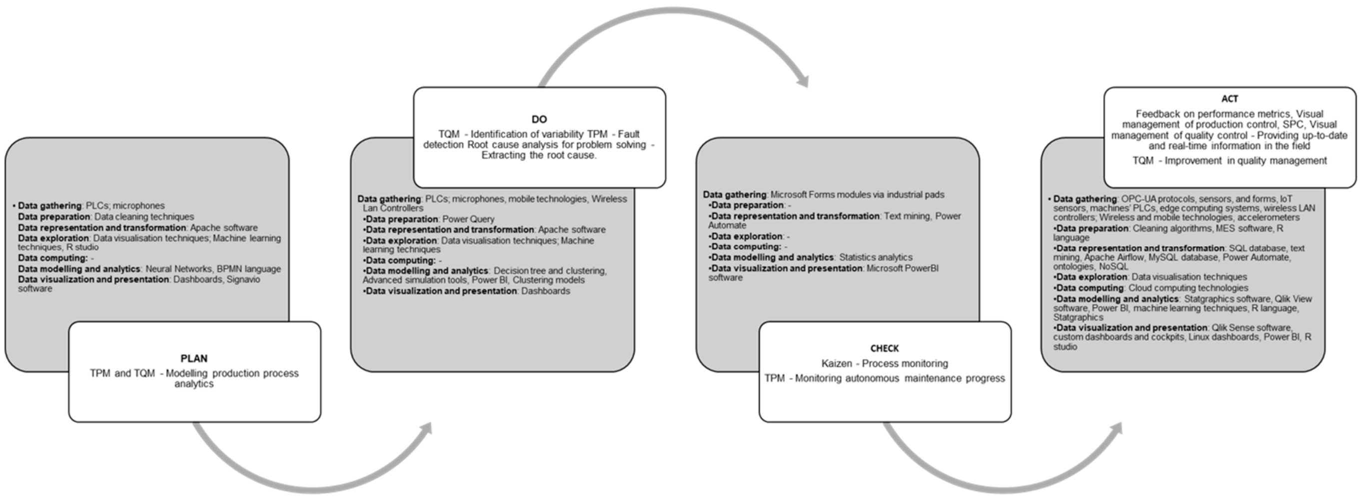

Figure 2. Summary of DS techniques and tools and LP practices from the cases, organised according to the stages of the PDCA cycle, adapted from [15].

Considering the LP practices that were already implemented by the companies before the introduction of DS, the techniques and tools contribute to supporting the check and the act stages. For example, in TPM, DS primarily boosts the check stage, providing data collection and an analysis of progress for process monitoring, whereas in TQM, DS primarily supports the act stage, providing a means to achieve process improvement with new information, such as the relationships between variables, or up-to-date and real-time information in the field.

Additionally, raised opportunities for the novel development of LP practices bring leveraging activities typical of the plan and do stages using DS. For example, in TPM, modelling production process analytics through DS supports the plan stage and allows the derivation of process models from event data and fault detection through machine learning techniques, developing predictive maintenance activities, i.e., the do stage. Conversely, in TQM, modelling production process analytics provides an alternative and insightful way to obtain the process model needed to plan the implementation of the LP practice, i.e., the plan stage, while the conduction of the process employing DS allows the identification of variability, i.e., the do stage.

6. Conclusions

In the context of I40, DS is expected to play a fundamental role. Manufacturing companies that adopt LP should be given a guide that clarifies how to support the implementation of the practices through DS. Motivated by a relevant gap in the literature that reflects the failure to address manufacturing companies' needs, this study adopts the multiple case study methodology and includes six Italian manufacturing companies that implement DS to support LP.

The results obtained contribute to theory and practice by providing insights into the debate on the interaction between LP and I40. Particularly, this research contributes to a discussion on the technologies specific to DS. From a theoretical viewpoint, this work extends the set of LP practices supported by DS that were reported by previous studies and empirically validates the application of DS within the PDCA cycle. From a practical viewpoint, the reported cases can inspire practitioners to exploit the value of data to facilitate and enhance the LP practices to which they apply. The detailed results of the empirical implementations highlighting the various DS techniques and tools used for DS activities in the context of LP practices provide references and suggestions for all managers evaluating alternative DS techniques and tools. Moreover, practitioners can understand different needs to implement diverse DS activities based on the speed and data structures, characterising their specific context.

Adopting a multiple case study methodology causes the generalisability of our findings to be the main limitation. Even if the applied methodology ensures the reliability and validity of this research, the results of our study cannot be generalised to a broader population. Nevertheless, the insights from the cases can provide interesting points of view that future research could verify. Research based on a survey conducted among a consistent number of manufacturing companies to confirm the evidence emerging from the six exploratory cases is already underway. Moreover, the sample consisted of Italian manufacturing companies. Although Italy is an important manufacturing country in Europe and is significantly investing in digital technologies, future research should include a broader sample of companies from different countries to confirm the obtained results.

Author Contributions: Conceptualization, R.P., V.G.C. and T.R.; Methodology, R.P.; Validation, T.R.; Formal Analysis, V.G.C.; Investigation, R.P.; Resources, T.R.; Writing—Original Draft Preparation, R.P. and V.G.C.; Supervision, T.R.; Project Administration, R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Womack, J.P.; Jones, D.T. Beyond Toyota: How to root out waste and pursue perfection. *Harv. Bus. Rev.* **1996**, *74*, 140–151.
2. Peças, P.; Encarnação, J.; Gambôa, M.; Sampayo, M.; Jorge, D. Pdca 4.0: A new conceptual approach for continuous improvement in the industry 4.0 paradigm. *Appl. Sci.* **2021**, *11*, 7671. [CrossRef]
3. Basulo-Ribeiro, J.; Amorim, M.; Teixeira, L. How to accelerate digital transformation in companies with Lean Philosophy? Contributions based on a practical case. *Int. J. Ind. Eng. Manag.* **2023**, *14*, 94–104. [CrossRef]
4. Naciri, L.; Mouhib, Z.; Gallab, M.; Nali, M.; Abbou, R.; Kebe, A. Lean and industry 4.0: A leading harmony. *Procedia Comput. Sci.* **2022**, *200*, 394–406. [CrossRef]
5. Xu, L.D.; Xu, E.L.; Li, L. Industry 4.0: State of the art and future trends. *Int. J. Prod. Res.* **2018**, *56*, 2941–2962. [CrossRef]
6. Klingenberg, C.O.; Borges, M.A.V.; Antunes, J.A.V. Industry 4.0 as a data-driven paradigm: A systematic literature review on technologies. *J. Manuf. Technol. Manag.* **2021**, *32*, 570–592. [CrossRef]
7. Ivanov, D.; Tang, C.S.; Dolgui, A.; Battini, D.; Das, A. Researchers' perspectives on Industry 4.0: Multi-disciplinary analysis and opportunities for operations management. *Int. J. Prod. Res.* **2021**, *59*, 2055–2078. [CrossRef]
8. Abd Rahman, M.S.B.; Mohamad, E.; Abdul Rahman, A.A. Bin Development of IoT—Enabled data analytics enhance decision support system for lean manufacturing process improvement. *Concurr. Eng.* **2021**, *29*, 208–220. [CrossRef]
9. Valamede, L.S.; Akkari, A.C.S. Lean 4.0: A new holistic approach for the integration of lean manufacturing tools and digital technologies. *Int. J. Math. Eng. Manag. Sci.* **2020**, *5*, 851–868. [CrossRef]
10. Tortorella, G.; Sawhney, R.; Jurburg, D.; de Paula, I.C.; Tlapa, D.; Thurer, M. Towards the proposition of a lean automation framework: Integrating industry 4.0 into lean production. *J. Manuf. Technol. Manag.* **2020**, *32*, 593–620. [CrossRef]
11. Anosike, A.; Alafropatis, K.; Garza-Reyes, J.A.; Kumar, A.; Luthra, S.; Rocha-Lona, L. Lean manufacturing and internet of things—A synergetic or antagonist relationship? *Comput. Ind.* **2021**, *129*, 103464. [CrossRef]
12. Agostini, L.; Filippini, R. Organizational and managerial challenges in the path toward Industry 4.0. *Eur. J. Innov. Manag.* **2019**, *22*, 406–421. [CrossRef]
13. Bittencourt, V.L.; Alves, A.C.; Leão, C.P. Industry 4.0 triggered by Lean Thinking: Insights from a systematic literature review. *Int. J. Prod. Res.* **2021**, *59*, 1496–1510. [CrossRef]
14. Ojha, R. Lean in industry 4.0 is accelerating manufacturing excellence—A DEMATEL analysis. *TQM J.* **2023**, *35*, 597–614. [CrossRef]
15. Cifone, F.D.; Hoberg, K.; Holweg, M.; Staudacher, A.P. 'Lean 4.0': How can digital technologies support lean practices? *Int. J. Prod. Econ.* **2021**, *241*, 108258. [CrossRef]
16. Pozzi, R.; Cannas, V.G.; Ciano, M.P. Linking data science to lean production: A model to support lean practices. *Int. J. Prod. Res.* **2021**, *60*, 6866–6887. [CrossRef]
17. Komkowski, T.; Antony, J.; Garza-Reyes, J.A.; Tortorella, G.L.; Pongboonchai-Empl, T. The integration of Industry 4.0 and Lean Management: A systematic review and constituting elements perspective. *Total Qual. Manag. Bus. Excell.* **2022**, *34*, 1052–1069. [CrossRef]
18. Buer, S.-V.; Strandhagen, J.O.; Chan, F.T.S. The link between Industry 4.0 and lean manufacturing: Mapping current research and establishing a research agenda. *Int. J. Prod. Res.* **2018**, *56*, 2924–2940. [CrossRef]

19. Rosin, F.; Forget, P.; Lamouri, S.; Pellerin, R. Impacts of Industry 4.0 technologies on Lean principles. *Int. J. Prod. Res.* **2020**, *58*, 1644–1661. [CrossRef]
20. Buer, S.V.; Semini, M.; Strandhagen, J.O.; Sgarbossa, F. The complementary effect of lean manufacturing and digitalisation on operational performance. *Int. J. Prod. Res.* **2021**, *59*, 1976–1992. [CrossRef]
21. Voss, C.; Tsikriktsis, N.; Frohlich, M. Case research in operations management. *Int. J. Oper. Prod. Manag.* **2002**, *22*, 195–219. [CrossRef]
22. Bai, C.; Satir, A.; Sarkis, J. Investing in lean manufacturing practices: An environmental and operational perspective. *Int. J. Prod. Res.* **2019**, *57*, 1037–1051. [CrossRef]
23. Watson, G.H.; DeYong, C.F. Design for six sigma: Caveat emptor. *Int. J. Lean Six Sigma* **2010**, *1*, 66–84. [CrossRef]
24. Yuik, C.J.; Puvanasvaran, P. Development of Lean Manufacturing Implementation Framework in Machinery and Equipment SMEs. *Int. J. Ind. Eng. Manag.* **2020**, *11*, 157–169. [CrossRef]
25. Eisenhardt, K.M.; Graebner, M.E. Theory Building from Cases: Opportunities and Challenges Linked references are available on JSTOR for this article: Theory Building from Cases: Opportunities and Challenges. *Acad. Manag. J.* **2007**, *50*, 25–32. [CrossRef]
26. Yin, R.K. *Case study Research and Applications*; Des. Methods; Sage Publication, Inc.: Los Angeles, CA, USA, 2018.
27. Classification Ateco (Classification of Economic Activity). 2007. Available online: https://www.istat.it/en/archivio/17959 (accessed on 7 March 2024).
28. Aida Analisi Informatizzata Delle Aziende Italiane. Available online: https://login.bvdinfo.com/R0/AidaNeo (accessed on 7 March 2024).
29. What Is an SME? Available online: https://web.archive.org/web/20150208090338/http:/ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index_en.htm (accessed on 7 March 2024).
30. Hays, P.A. Case study research. In *Foundations for Research: Methods of Inquiry in Education and the Social Sciences*; de Marrais, K., Lapan, S.D., Eds.; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 2004; pp. 217–234.
31. Cheng, Y.; Chen, K.; Sun, H.; Zhang, Y.; Tao, F. Data and knowledge mining with big data towards smart production. *J. Ind. Inf. Integr.* **2018**, *9*, 1–13. [CrossRef]
32. Boyer, K.K.K.; McDermott, C. Strategic consensus in operations strategy. *J. Oper. Manag.* **1999**, *17*, 289–305. [CrossRef]
33. Voss, C.; Johnson, M.; Godsell, J. 5 Case research. In *Research Methods for Operations Management*; Routledge: London, UK, 2016; Volume 165.
34. Stojanovic, L.; Dinic, M.; Stojanovic, N.; Stojadinovic, A. Big-data-driven anomaly detection in industry (4.0): An approach and a case study. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 1647–1652. [CrossRef]
35. Shahin, M.; Chen, F.F.; Bouzary, H.; Krishnaiyer, K. Integration of Lean Practices and Industry 4.0 Tech- nologies: Smart Manufacturing for Next-Generation Enter-prises. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 2927–2936. [CrossRef]