






Article

Trends and Challenges towards Effective Data-Driven Decision Making in UK Small and Medium-Sized Enterprises: Case Studies and Lessons Learnt from the Analysis of 85 Small and Medium-Sized Enterprises

Abdel-Rahman H. Tawil ^{1,†}, Muhidin Mohamed ^{2,*,†} , Xavier Schmoor ¹  and Konstantinos Vlachos ¹ 
and Diana Haidar ¹

¹ School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK; dianahaidar_91@hotmail.com (D.H.)

² Department of Operations and Information Management, ABS, Aston University, Birmingham B4 7ET, UK

* Correspondence: m.mohamed10@aston.ac.uk

† These authors contributed equally to this work.

Abstract: The adoption of data science brings vast benefits to Small and Medium-sized Enterprises (SMEs) including business productivity, economic growth, innovation and job creation. Data science can support SMEs to optimise production processes, anticipate customers' needs, predict machinery failures and deliver efficient smart services. Businesses can also harness the power of artificial intelligence (AI) and big data, and the smart use of digital technologies to enhance productivity and performance, paving the way for innovation. However, integrating data science decisions into an SME requires both skills and IT investments. In most cases, such expenses are beyond the means of SMEs due to their limited resources and restricted access to financing. This paper presents trends and challenges towards effective data-driven decision making for organisations based on a 3-year long study which covered more than 85 UK SMEs, mostly from the West Midlands region of England. In particular, this study attempts to find answers to several key research questions around data science and AI adoption among UK SMEs, and the advantages of digitalisation and data-driven decision making, as well as the challenges hindering their effective utilisation of these technologies. We also present two case studies that demonstrate the potential of digitisation and data science, and use these as examples to unveil challenges and showcase the wealth of currently available opportunities for SMEs.

Keywords: big data; data analytics; digitalisation; UK small and medium-sized enterprises (SMEs); AI adoption



Citation: Tawil, A.-R.H.; Mohamed, M.; Schmoor, X.; Vlachos, K.; Haidar, D. Trends and Challenges towards Effective Data-Driven Decision Making in UK Small and Medium-Sized Enterprises: Case Studies and Lessons Learnt from the Analysis of 85 Small and Medium-Sized Enterprises. *Big Data Cogn. Comput.* **2024**, *8*, 79. <https://doi.org/10.3390/bdcc8070079>

Academic Editors: Celestine Iwendi and Vishnu S. Pendyala

Received: 9 April 2024

Revised: 1 July 2024

Accepted: 4 July 2024

Published: 12 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2023, the UK reported a total of 5.6 million private sector businesses, marking a decrease of 7.1% compared to 2020. Small businesses, defined as those with 0–49 employees, constituted 99.2% of all businesses but contributed only 35.6% to the total turnover. Meanwhile, Small and Medium-sized Enterprises (SMEs), encompassing businesses with 0–250 employees, represented 99.9% of UK businesses and generated 52.5% of the total turnover. Notably, the number of non-employing businesses decreased by 10% between 2020 and 2023, while employing businesses saw a modest increase of 2.3%. At the start of 2024, the average turnover of all UK businesses rose by 6.9% compared to 2022, amounting to £806,381 [1]. These data show the importance of SMEs for the UK economy and how powerful any steps taken to assist their rapid growth would be in boosting the economy of the country. Between 2020 and 2021, as a consequence of the COVID-19 pandemic and lockdown measures, the number of businesses in the UK decreased by 6.5%. SME numbers

fell across all regions and countries in the UK—the greatest fall occurred in Northern Ireland, where businesses fell by 16.6%, followed by London by 8.0% and Scotland by 7.4% [2]. Moreover, in emerging economies, SMEs are estimated to generate 60% of employment and 40% of Gross Domestic Product (GDP), while in the European Union the proportion of the workforce employed by SMEs is higher, 66% [3].

SMEs and their investors are recognising the value data provide for their business [4]. Contemporary companies worldwide [5–8] and typically in the UK seek data-driven innovations not only to modernise business operations and increase their competitiveness advantage, but also to carve out new markets, and meet varying government policies and numerous regulators, as well as make their businesses more sustainable [9]. An IBM report cited in [10] states that 2.5 quintillion bytes of data are generated every day. Remarkably, 90 percent of the world's data has been created in just the past decade, making it the “new oil” of this digital era [11]. Data are like crude oil, without analysis they are of little use to businesses if they do not know how to process or use them. Technologically efficient companies are among those that achieve high growth rates, according to research [12,13]. Data-drivenness is about building tools, abilities and more crucially a culture that acts on data. A leading factor that shapes this transformation is the data collected in databases and other repositories maintained by the business. Companies that see data as a strategic asset will thrive as data becomes a key part of their competitive advantages in the coming years. Obviously, not just any data will work; they have to be the right data (e.g., timely, accurate, clean, unbiased and most importantly trustworthy). Good data have the power to transform businesses with the actionable insights required to become more productive. Evidently, there can be subtle hidden biases in the data that can sway drawing the right conclusions. However, cleaning and managing data can be tough, time-consuming and expensive operations.

Data scientists use data analysis techniques to develop new business models that are used to deliver, create and capture value for business growth, success and profitability. Their skills are now essential to industry transformation. The analytic value chain in a data-driven organisation stimulates deeper analysis. Decision makers usually incorporate these into their decision-making processes so they can influence the direction taken by the company, and therefore add value and impact. This process transforms data into knowledge and value, which creates new income streams. Yet, despite the benefits and opportunities digital technologies bring, and despite the significant uptake in recent years, many SMEs are still lagging in the adoption of digital technology and, for smaller SMEs with 10–49 employees, the digital adoption gap has widened significantly compared to larger firms [14]. For example, SMEs in the UK are adopting big data analytics at a rate of less than 1% [15,16]. However, recently businesses in the UK are becoming more aware of the value of data-driven decision making and data analytics is increasing in popularity, according to recent reports. The data science industry is also poised to expand over the next few years.

This paper aims to explore trends and challenges towards effective data-driven decision making for UK businesses, how SMEs pivot their business models around data to handle data-driven products, and how this contributes to their innovation and performance. We present an analysis of the challenges and opportunities of digitalisation, and adoption of data science and artificial intelligence (AI) within the UK SME business sector. Our analysis of 85 UK SMEs is based on case studies of SMEs located primarily in England's West Midlands who are supported in the areas of data management, machine learning, data analytics and other related digital technologies under a 3-year long European Regional Development Fund (ERDF) project named Big Data Corridor (BDC). Our study also briefly examines how small businesses can take advantage of data-driven innovation and decision making, while highlighting challenging areas where support for digital technology adoption is most needed.

The multi-perspective analysis and case studies in this paper inform the SME business industry as well as business innovation and growth bodies about potential challenges and

key opportunities in AI usage. In addition, the analysis encourages small businesses to derive meaningful insights from closed (private business) and open (publicly available) data by taking advantage of emerging data science and AI techniques and technologies. The research also highlights areas where future support and funding are most needed to enable SMEs to embark on the digital revolution and AI adoption thereby contributing to the growth and development of this key sector in the UK.

This study focuses on the lessons learnt from the analysis of business data and the adoption of data-driven solutions by SMEs. The specific contribution of this paper is to find answers to the following research questions:

1. How can data science and AI adoption benefit Small and Medium-sized Enterprises (SMEs) in terms of business productivity, economic growth, innovation and job creation?
2. What are the challenges faced by SMEs in integrating data science decisions into their operations, considering limited resources and restricted access to financing?
3. What are the potential benefits of data analytics and digital transformation for SMEs, including marketing optimisation, demand forecasting, and customer retention and acquisition?
4. How can SMEs in the UK take advantage of data-driven innovation and decision making, and what areas require the most support for digital technology adoption?

The rest of the paper is organised as follows. Sections 2 and 3 present related works and the research methodology used in order. In Section 4, we analyse SMEs' data, digital technology trends, faced challenges, and the key lessons learned while supporting and collaborating with businesses. Section 5 covers two case studies on SMEs that embarked on digitisation and data science adoption. We provide a summary and conclusion in Section 6.

2. Related Works

Data analytics and digital transformation offer businesses new opportunities, such as marketing optimisation, the forecasting of demand for their products and services, and staying one step ahead in retaining and acquiring customers. In a related study, Bhardwaj [17] provides a comprehensive review of 42 peer-reviewed studies from 2010 to 2021 on data analytics in SMEs. The review identifies four main themes: enabling factors, restraining factors, investing SMEs and performance indicators. It highlights the significant role of data analytics in enhancing SME competitiveness and identifies barriers such as poor IT infrastructure and lack of analytics knowledge. The paper emphasises the need for more research on underexplored themes and suggests future research directions to bridge existing gaps. This work consolidates current knowledge and guides future studies to improve the strategic use of data analytics in SMEs. Also, a survey of 500 UK companies found a positive correlation between the use of data and business performance and productivity: top data-using companies are 13% more productive than those in the lowest quartile [18]. Many government institutions, including the EU, recognise the importance of empowering SMEs to benefit from the digital revolution and generate measurable economic benefits. This is evident from the proportion of EU funding allocated to data-related projects, big data and data science [19]. To increase the number of highly skilled workers in AI and data science, the UK government, the Office for Students, universities and industry partners have established a fund of up to £24 million [20].

In another related work, Schönberger [21] examines the adoption of artificial intelligence (AI) by Small and Medium-sized Enterprises (SMEs), highlighting key applications, benefits and challenges. The study employs a quantitative research approach through an online survey distributed among German SMEs, focusing on AI tools like virtual assistants, recommendation systems and machine learning. The findings reveal that these technologies enhance efficiency, productivity and decision making, but also present challenges such as privacy concerns and the need for specialised skills. Despite limited resources hindering AI adoption, the study underscores the potential of AI to transform business processes in SMEs and serves as a basis for future research and practical guidance for SMEs considering AI

implementation. Furthermore, Griesch, Rittelmeyer and Sandkuhl explore AI-as-a-Service (AIaaS), which leverages AI and cloud computing to provide accessible AI solutions for Small and Medium-sized Enterprises (SMEs) [22]. The paper addresses the research gap concerning the differences between AIaaS and on-premise AI implementations. It includes a literature review to identify factors affecting AI adoption and a detailed case study comparing AIaaS with on-premise AI in a real-world SME context. The study also employs a morphological box to systematically compare these approaches, highlighting AIaaS's potential to overcome SMEs' technical and resource limitations while detailing its practical applications and limitations.

Our paper sheds light on the research questions set out in Section 1 by analysing data from 85 SMEs in the West Midlands region, focusing on their digitisation trends, challenges faced and lessons learned from adopting data-driven solutions. The two case studies presented in the paper (Section 5) will serve as examples to demonstrate the potential benefits and challenges of implementing digitisation and AI in SMEs. By addressing these research questions, this paper seeks to contribute valuable insights to the SME business industry and encourage the growth and development of this sector in the United Kingdom through the effective use of data and emerging trends in machine learning and analytics.

3. Research Methodology

The objective of the research work presented in this paper is to determine and analyse the needs, lessons, challenges and opportunities of the UK SMEs to digitise their business processes and adopt AI and data-driven methods. The findings reported in this work are based on the 3-year long ERDF project, BDC. In particular, the study explores digital technology trends, challenges limiting SMEs in effective utilisation of enabling technologies and the state of their adoption in data science technologies. Business opportunities and advantages of digitisation and adopting some data analytic technologies and AI are demonstrated through two selected practical case studies in Section 5.

The overall research framework employed in this study is illustrated in Figure 1. In the first step, SMEs were recruited for the BDC project (cf. demonstration in Figure 2). Stages 2 and 3 of the research methodology are concerned with SMEs' data collection and analysis in order to identify their digital technology trends and challenges to embarking on the route to digitalisation and use of data-driven methods. In stage 4, specific case studies were developed to respectively demonstrate digital transformation and AI use by SMEs.

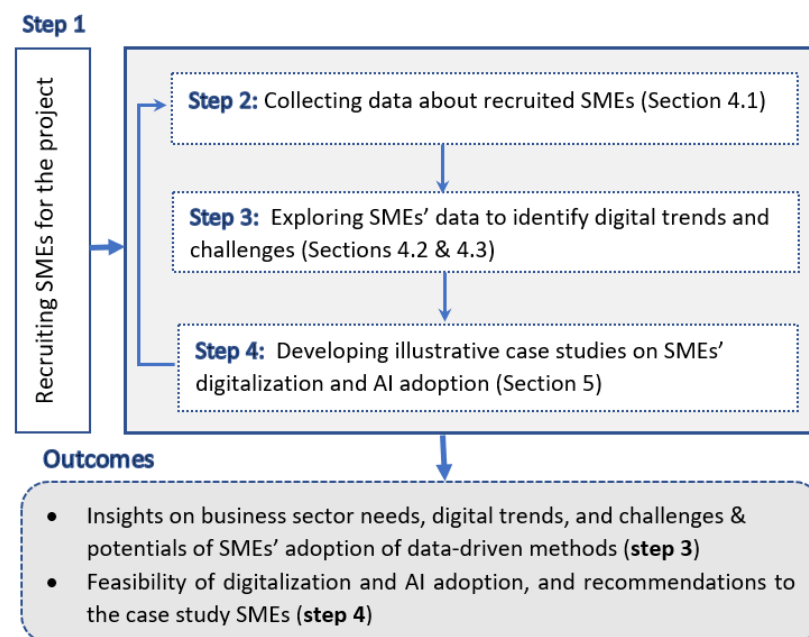


Figure 1. Research methodology and outcomes.

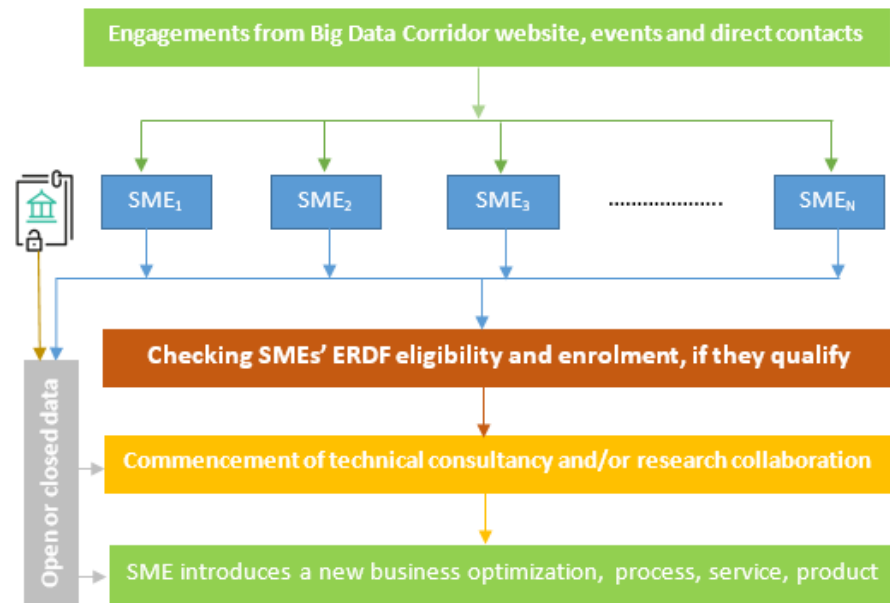


Figure 2. BDC project workflow for supporting SMEs.

4. Analysis of SMEs' Digitisation Trends, Challenges and Learned Lessons

4.1. Data Collection

Eighty-five SMEs received support during the period from June 2017 to August 2019 from two academic partners of the BDC project and are used as the basis for our analysis and discussions. Figure 2 depicts the BDC project support life cycle for SMEs including SME engagement, checking SME eligibility and commencement of support which could lead to the introduction of new products or services. Once recruited, SMEs' data are primarily collected via designed project forms and structured meetings with them. These SMEs were mostly based in the West Midlands region of the UK and more precisely in the Greater Birmingham and Solihull Local Enterprise Partnership (GBSLEP) area, which was the focus of the BDC project. Due to non-disclosure agreements and the need to preserve business confidentiality, we do not detail individual SMEs' support or provide any associated private data except in the two case studies that were used as examples with the permission of these businesses, which are also anonymised.

4.2. SMEs' Digitisation Trends from Different Perspectives

The sample SMEs had a minimum of one (the owner) and a maximum of 146 employees. The minimum turnover of the SMEs was £10,000 and the maximum reported turnover was £10 million. Companies were categorised based on their activities into 20 different sectors. Table 1 shows the top 10 business sectors based on the count of SMEs, together with ranges for the number of employees and turnover for the SMEs of each sector.

The top 10 business sectors presented in Table 1 together account for 78.9% of the total SMEs from the sample. Notably, most of the SMEs were from the Information and Communications Technology sector with 16 businesses representing 18.8% of the total companies involved in the project. Other sectors represented with a higher number of SMEs include Education and Training (10 SMEs and 11.8% of total), Consultancy (10 SMEs and 11.8% of total), Marketing and Public Relations (7 SMEs and 8.2% of total) and Human Health and Social Work activities with 6 SMEs and 7.1% of total. These numbers suggest that companies in the technology and services sectors rely more on data-driven solutions to support their business when compared to companies in other sectors.

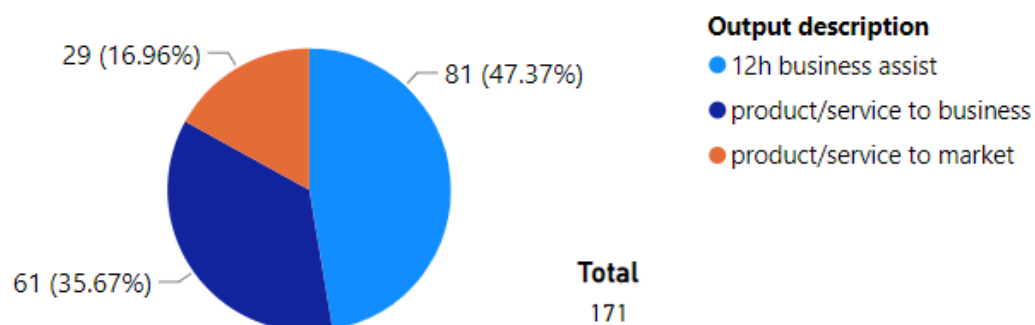
Table 1. Top ten business sectors among the 85 SMEs supported by BDC project.

Sector	SME Count	SME Proportion	No of Employees	Turnover
Information and Communications Technology	16	18.8%	1–71	£10 k–£7 m
Consultancy	10	11.8%	1–20	£10 k–£1.4 m
Education and Training	10	11.8%	1–70	£10 k–£6.5 m
Marketing and Public relations	7	8.2%	1–15	£10 k–£250 k
Human Health and Social Work Activities	6	7.1%	1–35	£10 k–£650 k
Travel and Tourism	5	5.9%	1–5	£10 k–£700 k
Manufacturing	4	4.7%	1–55	£10 k–£850 k
Accounting and Finance	3	3.5%	1–146	£10 k–£10 m
Arts, Entertainment and Recreation	3	3.5%	1–10	£10 k–£400 k
Food	3	3.5%	1–28	£10 k–£1.2 m

The SMEs included in this analysis completed at least one of the following types of business support (outputs) provided by the BDC project:

1. Twelve hours of business assist and training;
2. Support to introduce a new product/service to the business;
3. Support to introduce a new product/service to the market.

In Figure 3, the count and proportion of project outputs by type are shown.

**Figure 3.** Count and proportion of project outputs by type.

The total number of outputs was 171, which is higher than the total number of SMEs involved, indicating that some companies received multiple types of business support. The most common type of output was the 12-h business assist, with 81 assists in total. This type of support benefited the majority of companies (81 out of 85 or 95%), and included seminars, workshops and quick assistance for non-complex data-related business problems.

Among the remaining outputs that required longer-term collaboration, 61 SMEs introduced a new product or service internally, while 29 SMEs introduced a new product or service to the market. These results indicate that companies primarily sought data-related support to upgrade their internal services, gain insights from their data and improve decision-making processes. It is worth mentioning that some outputs not introduced to the market were either Proof of Concept (PoC) solutions or components of larger projects intended for future market release.

Figure 4 presents counts of support type by business sector.

Apparently, the sectors with the most sought technical support were Information and Communication Technology (32 outputs) and Education and Training (29 outputs). However, considering the count of SMEs from each sector, the Education and Training sector had almost three assistances per SME (29 outputs for 10 SMEs) while Information

and Communication Technology had two per SME (32 outputs for 16 SMEs). In this category, the Manufacturing sector comes first with 14 outputs for four SMEs (3.5 outputs per SME). The sector that received most of the 12-hour support was Information and Communication Technology with 14 outputs followed by the Education and Training and Consultancy sectors with 10 outputs each. The Information and Communication Technology sector also received most of the product/service to market support with 6 outputs, and the Education and Training sector received most of the product/service to business support with 14 outputs.

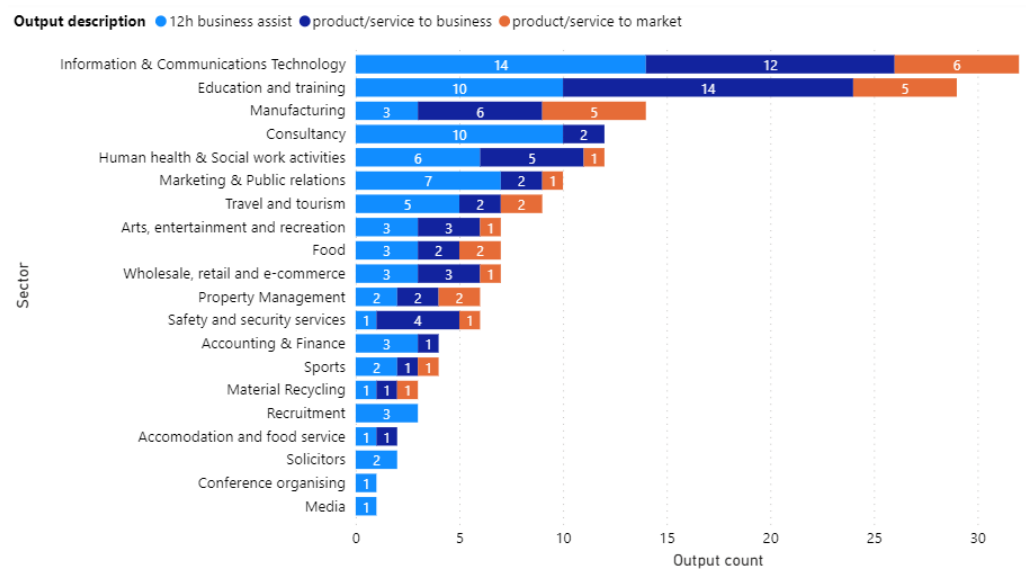


Figure 4. Count of project outputs per category by sector.

For the purpose of this analysis, we further introduced four categories of digital support based on the requirements of the support provided to businesses from the project’s academic partners:

1. Skills development and training: Companies attended training seminars and workshops focused on data collection, transformation and storage as well as seminars about data analysis, visualisation and using data science and machine learning to get better insight and make data-driven decisions.
2. Data management and analytics: Supported SMEs to collect and store data more efficiently, in order to make them accessible for analysis and visualisation and allow for better business insight extraction.
3. System design and development: Supported SMEs to design and develop end-to-end business solutions either to improve their products and services or to enhance internal business decision making and planning.
4. Other: This refers to a small number of SMEs provided with support that does not directly fall under one of the above three categories, such as recycling, fire and safety, etc.

As illustrated in Figure 5, 34 companies (40%) required support related to skills development and training, 29 (34.1%) for system design and development, and 19 (22.3%) for data management and analytics. Across all different types of support, the main data science and analytics aspects in which the companies required assistance can be summarised as follows:

1. Acquiring data (open or proprietary) from external web resources such as APIs, websites or web repositories, and transforming, filtering or combining these data to extract business insight.
2. Creating advanced visualisations including interactive dashboards and performing descriptive analytics to gain insight and improve products/services or marketing practices.

3. Collecting, storing and analysing data originating from sensors including IoT.
4. Using machine learning (ML) to improve their products/services with data-driven processes or to provide recommendations for their customers.

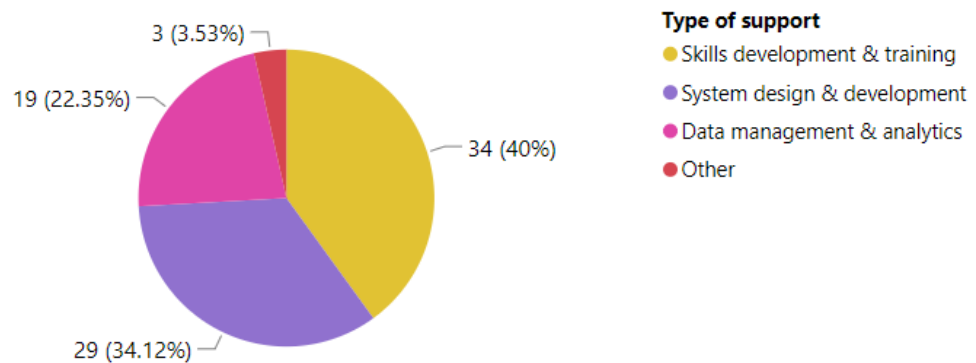


Figure 5. Count and percentage of SMEs per type of support received.

Figure 6 presents the number of outputs per type of support. It is evident that the project produced a higher number of outputs for system design and development and data management and analytics support when compared to skills development and training, especially for interventions that assisted companies to introduce new products and services internally or to the market.

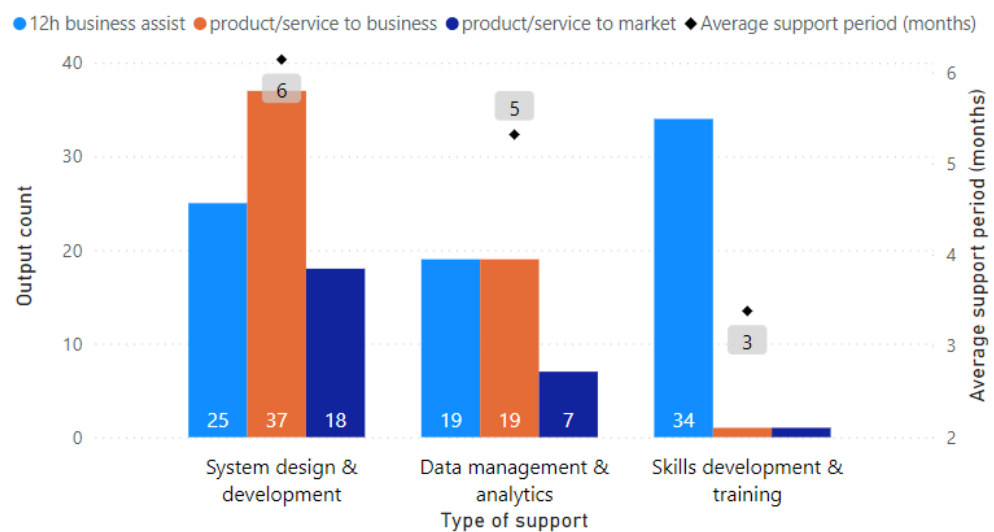


Figure 6. Output count by type and average support period per type of support.

The duration of support received by SMEs ranged from 1 to 16 months and the average support period was slightly below 5 months (4.75) per SME. As evident in Figure 6, companies that received support to design and develop data-related solutions or to improve their data management and analytics processes required longer support periods on average (5–6 months) when compared to companies that received support for skills development and training (average support period of 3 months). In Figure 7, we show the top 10 business sectors supported and the average support period received by SMEs from each sector. SMEs in the Education and Training sector had the highest average support period (8 months), followed by Accounting and Finance (7 months) and Human Health and Social Work Activities (6 months). In general, it was observed that longer support periods were required for sophisticated projects, either because these projects demanded expert knowledge, or due to lower levels of data science and technical skills within the companies.

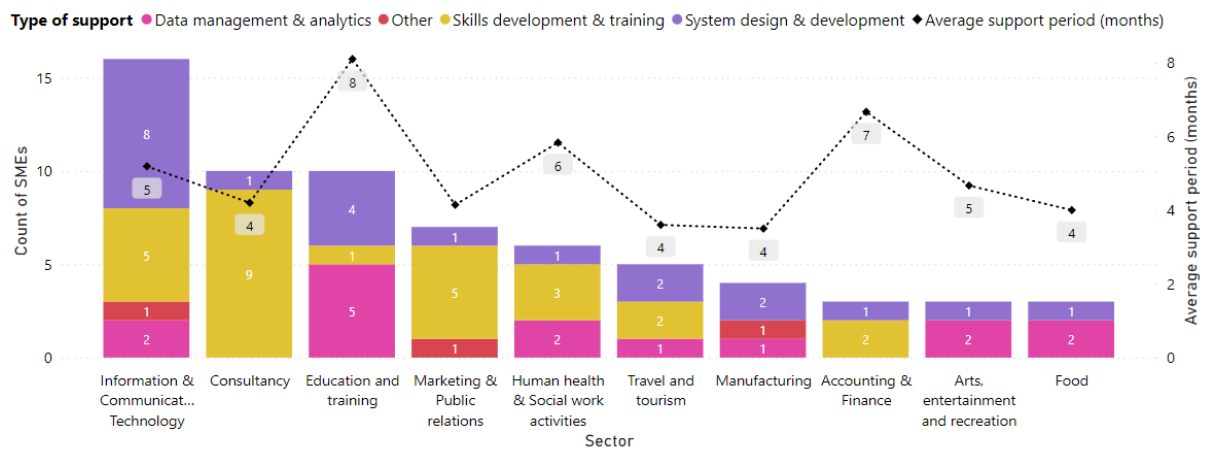


Figure 7. Count of SMEs receiving support by type of support and average support period per sector.

Another important observation derived from Figure 7 is that system design and development support was requested by companies across all sectors. This highlights the high demand for data-driven solutions and services in the market which corroborates with the main aim of the study, to encourage digital transformation and use of data in SMEs to achieve growth.

4.3. Challenges and Learned Lessons

The digital transformation of SMEs and their adoption of data science techniques require organisational changes, investment and resources before businesses can reap any resulting profits and benefits. However, we have observed during this study that the majority of micro and small enterprises are hesitant to invest in new digital tools and methods if they cannot anticipate quick positive results and revenues. From our analysis, we have empirically recognised four main areas (without excluding other potential ones) in which SMEs effectively utilise data-driven strategies and techniques to enhance their business operations.

- First, improving digital marketing proved to be a common data-driven use case among small businesses. Specifically, a large number of the study's SMEs managed to derive immense marketing insights from analysis performed on their own data or relevant open data, e.g., identifying lead customers, gaining a better understanding of customer purchasing patterns, etc.
- Second, the use of social media (SM) platforms for digital marketing was found popular among our case study SMEs. This is driven by the valuable data generated because of customers communicating with SMEs via their dedicated channels on these platforms. The SM analytics was used by some enterprises to study their marketing performance on various platforms (e.g., Instagram, Facebook, LinkedIn and Twitter) about customer purchase and service use intents. This leads SMEs to retain the highest revenue-generating platforms and unsubscribe from underperforming ones, eventually reducing costs. The same analytics was also used to identify working marketing strategies based on specifically studied Key Performance Indicators (KPIs), e.g., views, likes, reach, impressions, etc. Table 1 and Figure 7 show that Marketing and Public Relations represent the top support-seeking business sectors according to our study. Linked to this, a study on the adoption of social media and SMEs' performance found that around 70% of SMEs use social media for marketing [23].
- Third, a good number of the SMEs supported under BDC used data-driven methods to identify potential new markets by performing cross-analysis of their customer data and various relevant open datasets, e.g., government open datasets, such as census, demographics and geospatial data. Such customer analytics enabled SMEs to locate

new markets with a high demand for their services and products, thus allowing them to expand their market presence.

- Eventually, performing exploratory analysis on historical data with the objective of creating predictive models was another common application observed among the case study SMEs. Going to the predictive stage of the analytics ladder enables many small and medium-sized businesses to make use of cutting edge ML and AI technologies. The selected two case studies presented in Section 5 represent good examples for this type of application. They include an SME transforming from manual to digital business operations and another adopting machine learning for parameter optimisation in additive manufacturing.

Linked with the above, the research and consultancy work with SMEs under this project has also produced other lessons. In particular, we underline a number of factors found to be the leading barriers preventing most SMEs from embracing new and emerging data-driven technologies and AI. Figure 8 summarises the primary challenges facing SMEs to successfully adopt digitisation and data-driven approaches as identified in this study. The challenges are roughly classified into three broad categories: data-related, management and financial, in addition to expertise and knowledge.

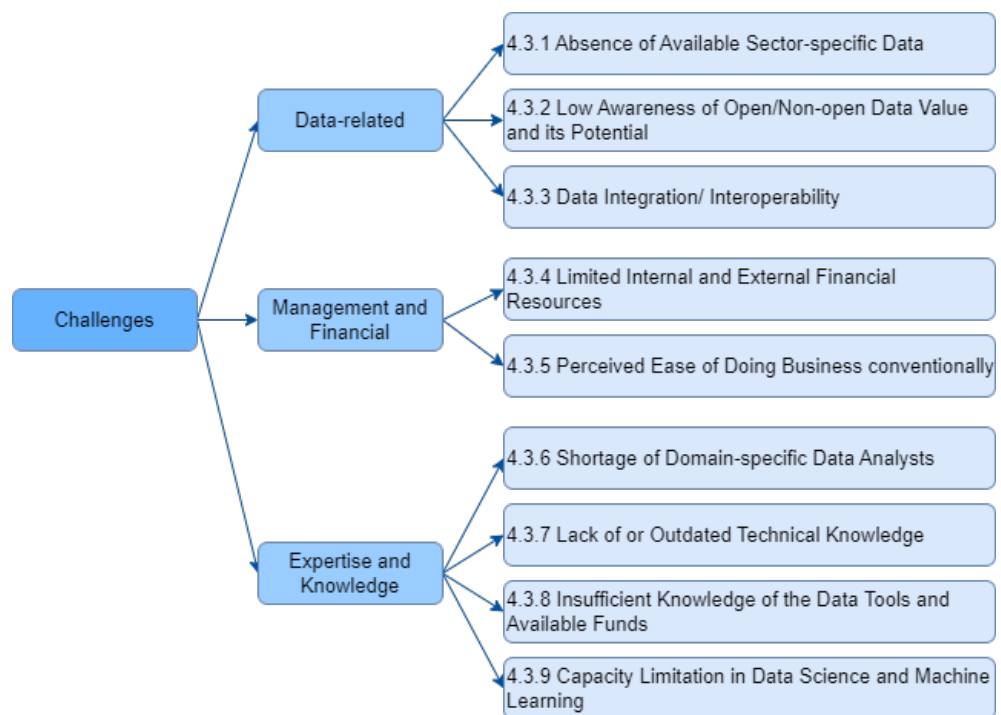


Figure 8. Identified challenges facing SMEs in adopting data-driven approaches.

4.3.1. Absence of Available Sector-Specific Data

Some AI and big data analytics applications are relatively new in the context of their deployment and use for some specific domains. This means that any meaningful and successful application in such areas will be hampered by the lack of existing analytic data. For example, we performed a pilot study for one SME by applying supervised machine learning to input–output parameter optimisation in additive manufacturing, and one main limitation we found was the insufficiency of fabrication data in the 3D printing sector. This is not to suggest that there is no reasonable closed data among the SMEs in this sector, but that may not be enough to implement supervised machine learning models and there are no feasible current mechanisms by which such SMEs can share their own hard-earned limited data. This led to suggestions of using simulated data for ML-based parameter optimisation in AM [24].

4.3.2. Low Awareness of Open/Non-Open Data Value and Its Potential

In many cases, government datasets are made available for free and are made public through data openness. Usually, the reasons are related to improving the service, increasing economic value or achieving political transparency. These datasets can be valuable to the success of small businesses, but they seem to be largely unknown to them. Around 20 percent of the SMEs we supported created new services and discovered new markets using open data from education, e-recruitment and transportation. The more SMEs understand open data advantages and how it can deliver positive outcomes, the more innovative products and services they are able to design using it at a lower cost. Moreover, many of the SMEs we worked with were aware that they could use data not only to keep records but also to gain useful insights into their business. Yet, learning data science skills is a challenge for SMEs as many are unsure of the types of data from their operations that should be collected. The fact that SMEs lacked knowledge about data handling tools also made them unable to use this data.

4.3.3. Data Integration/Interoperability

The competitiveness of SMEs can be improved by integrating their systems with their suppliers or other open data available from governments or trading partners [25]. The demand for businesses to exchange real-time information has grown since the adoption of technologies such as mobile commerce, electronic funds transfers, supply chain management and online transaction processing. Organisations could benefit from integrating their IT infrastructures to address this need. In many cases, however, portability and interoperability measures for SMEs are still being implemented relatively slowly, while in others they are at an early stage. Data integration is a significant problem for SMEs due to high costs and technological requirements and is frequently cited as being a sticking point within organisations. Some existing approaches to integration, such as Electronic Data Interchange (EDI), inventory management and automated data collection systems, can help SMEs overcome some of their integration challenges, but they have their limitations. It is possible to further simplify these integration problems by utilising web services nowadays. The need for case studies on the integration efforts of SMEs and how they deal with integration and interoperability problems is critical.

4.3.4. Limited Internal and External Financial Resources

In our research work with SMEs, we discovered that most of them are able to recognise and utilise data and analytics to grow their businesses. Also, we noted their willingness to use data analytics and invest in related technologies to derive meaning from their data. Despite their attempts at adopting data technology, their capabilities are insufficient since most do not possess the required knowledge and expertise, nor the budget to hire specialists or outsource their analytic needs. A further barrier comes from SMEs having limited access to funding and loans compared to big companies. Thus, the significance of this ERDF and other similar projects, which aim to bridge the gap, is justified. A greater level of financial and technical support is also needed to foster the adoption of data analytics by micro- and small-sized businesses, especially during the period of COVID-19 [14].

4.3.5. Perceived Ease of Doing Business Conventionally

The majority of SMEs operate in specific business sectors and are dependent upon conventional business methods. Thus, many business owners would prefer to maintain their traditional business practices and refrain from adopting disruptive technologies, such as artificial intelligence, data analytics, etc. As a consequence, they are not tempted to use data other than for record-keeping purposes. There are a number of possible explanations for this reluctance to leverage the benefits of advanced data usage—not just a misunderstanding of what these advances are meant to offer their companies, but also the perceived short-term disruption such changes might cause—such as learning new tools, implementing software packages or investing in cloud computing, and hiring employees

with data analytics skills. However, the result is an overlooked data-driven business opportunity, which could have been beneficial for the SMEs in the long run. SMEs' decision-making process could be made more data-driven by encouraging a business shift.

4.3.6. Shortage of Domain-Specific Data Analysts

It is often necessary to combine data analytics skills with business context and domain knowledge in order to analyze business data effectively [26]. As a result of the limited number of analysts in the business who meet such criteria, SMEs outsource their data analytics needs to experts at comparatively higher costs. Data science and machine learning for process control and modelling are still in their infancy in some domains, e.g., additive manufacturing, due to the lack of many experts who understand the business context. SMEs in manufacturing are less likely to use data for analytics and decision-making [27], which makes it harder for them to exploit their data to enhance growth and productivity.

4.3.7. Lack of or Outdated Technical Knowledge

Small businesses need to develop and upgrade employee skills in order to grow. The lack of finance, capacity in data science and machine learning, and awareness of innovation funds, among other things, prevent many SMEs from remaining competitive as technology advances in terms of employee skills and competence. These are especially applicable to smaller companies, such as manufacturing and technology companies, who face challenges when designing, developing and testing new products, or upgrading their existing ones to meet new needs. Based on the analysis in Section 4.2, we can understand that the percentage of SMEs with skills and training development needs is roughly 40% as shown in Figure 5.

4.3.8. Insufficient Knowledge of the Data Tools and Available Funds

In our study, we found that SMEs are unaware of the available funding, especially for adopting innovative digital and data-driven ideas. Knowledge Transfer Partnerships (KTPs) are considered one of the most appealing funding schemes with respect to the BDC's digital assistance and collaboration with SME business owners since minimal contributions are required from companies and the partnership may have a high level of impact on SMEs. In addition, our study found that only a very small percentage of micro enterprises are aware of some of the most widely used data management and analysis tools, and the benefits these tools can provide. It proved crucial to many SMEs to support their adoption of these tools (including Dropbox, Power BI and R Studio).

4.3.9. Capacity Limitation in Data Science and Machine Learning

One other primary challenge facing SMEs in digitising and adopting AI and data science technologies is the inadequacy or lack of working knowledge and capacity in these technologies. This barrier is partially linked to the other aforementioned challenges such as SMEs' limited resources, e.g., finance. Figure 5 shows the breakdown analysis of the digital technology support sought by the sample SMEs of which skills development and training forms the largest percentage (40%) followed by system design and development (34%). For some SMEs, this was partially compensated by outsourcing data science and AI needs to external solutions such as cloud computing-based Machine Learning as a Service (MLaaS) platforms [14].

5. Case Studies in Digitisation and Adoption of AI and Analytics in SMEs

In this section, we present two illustrative case studies that were conducted under the research project. The first one (Section 5.1) is about the digitisation journey of an SME, from using paper forms for business operations to using digital solutions. The second case study (Section 5.2) looks at an additive manufacturing company that embarks on the route to AI and machine learning adoption for parameter optimisation.

5.1. A Data-Driven Solution for Monitoring the Delivery of PBL Care Services

PBL Care is a domiciliary care service SME based in Birmingham—West Midlands [28]. It provides home care and support to individuals who live independently in their own homes. The SME offers a wide range of services such as personal care, assistance with eating and toileting, medication support and palliative care. Among the individuals supported by PBL Care, some have physical disabilities, dementia and mental health conditions. PBL Care is regulated by the Care Quality Commission (CQC), part of the UK Department of Health and Social Care, responsible for the regulation and inspection of health and social care services in England [29].

5.1.1. Problem Statement

The CQC inspected PBL Care back in December 2017 and rated the SME as “Requiring improvement”, meaning that the service was not performing as well as expected. The CQC report highlighted that, at the time of the inspection, the company did not have the right processes in place to guarantee effective monitoring of the delivery of care, resulting in the late arrival of care staff at patient homes and most visits not lasting as long as planned. In addition to this, most activity records were paper based. The literature describes that digitisation of patient records can improve communication and coordination in health care organisations [30,31], especially in the home-care context [32]. Accordingly, after several meetings took place with the SME to understand their requirements, it was agreed to digitise the business in order to meet evolving demands and keep pace with the rapidly changing home-care sector.

5.1.2. Methodology

The case study aims to create a data-driven solution for PBL Care that reflects the implementation of new processes and allows the SME to digitally monitor the delivery of care. In order to achieve this, the following work plan was set:

- Step 1: Identification of Key Performance Indicators (KPIs);
- Step 2: Data collection;
- Step 3: Data visualisation and dashboard prototyping;
- Step 4: Data interpretation and evaluation.

The aim of the last step is to analyse outcomes following the implementation of new processes by the PBL Care management team. In the following sections, we will describe each of these steps.

5.1.3. Development of a Digital Solution

Identification of KPIs

In order to identify KPIs, we started by looking at paper-based information collected by PBL Care. The SME used paper “log books” to record information such as date, time in, time out, carer’s name, food/fluids taken, pad changed, tasks carried out, concerns raised, etc. Care staff fill one log book per visit at the service user’s (patient’s) location. We identified the following attributes as key for monitoring: *Patient ID, Care staff ID, Date, Time In and Time Out*.

Data Collection

Data collection templates were created using Excel spreadsheets to digitally gather information based on the identified KPIs. PBL Care staff were then trained to enter data in the spreadsheets. For the first iteration of the data collection and dashboard prototyping, this was completed in two steps: first PBL Care staff filled in paper log books when visiting patients at their homes, then other PBL Care staff entered the collected data into Excel spreadsheets from the PBL Care office. This process has been improved for the next iterations of data collection and dashboard prototyping, as discussed in Section 5.1.4.

Data Visualisation and Dashboard Prototyping

Guidance from the National Institute for Health and Care Excellence (NICE) states that home-care visits to elderly people should last for at least half an hour unless specific circumstances are met [33]. Based on this, we built an interactive dashboard using Microsoft Power BI (Version: 2.6) to monitor the duration of visits as presented in Figure 9. The dashboard enables users to select a month, a service user (patient), a care staff member or a histogram range by clicking on an ID or time range. The information displayed on the dashboard is filtered interactively based on the user selection. The dashboard allows the PBL Care management team to look for specific service users or care staff, and to check whether NICE guidance on visit duration is followed. Users can get the latest information entered in the data collection spreadsheets by refreshing the dashboard in one click, making it easy for PBL Care to visualise the latest data.

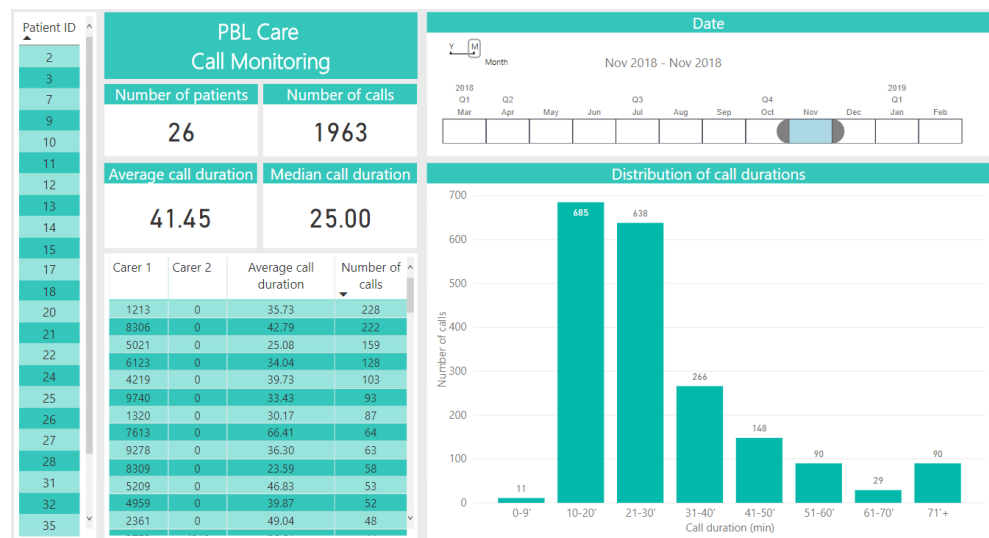


Figure 9. Dashboard prototype for monitoring of visits duration.

Data Interpretation and Evaluation

The histogram on Figure 9 highlights that most calls (visits made by care staff) made in November 2018 lasted for a duration between 10 and 20 min (685 calls), which is not good practice according to NICE guidance as mentioned in Section 5.1.3. This might be due to calls being scheduled back-to-back, not giving enough time for care staff to travel between patient homes and forcing them to shorten visits. PBL Care introduced two measures to address this problem: care staff schedules were adjusted to allow sufficient travel time between patient homes; and care staff were reminded of the importance of being on time for their visits and made aware that controls were carried out.

Data for the three consecutive months of November, December and January 2019 were collected to analyse the impact of the actions taken by PBL Care. Figure 10 presents the average visit duration, median visit duration, percentage of visits between 10 and 20 min, and percentage of visits between 8 and 10 h.

The average visit duration is around 50 min for December and January, which is an increase of 8 min from November. This could be due to PBL Care taking more night-shift NHS packages from December, as shown by the increase in percentages of visits between 8 and 10 h from 1.1% in November to 2.0% in December. The median visit duration, more robust with regard to extreme values, is 30 min for December and January. We can observe a consistent decrease in the percentage of visit durations lasting between 10 and 20 min, from 34.9% in November to 29.8% in December and finally 26.1% in January. This decrease could be linked to the two actions taken by the PBL Care management team and a sign of improvement in care delivery.

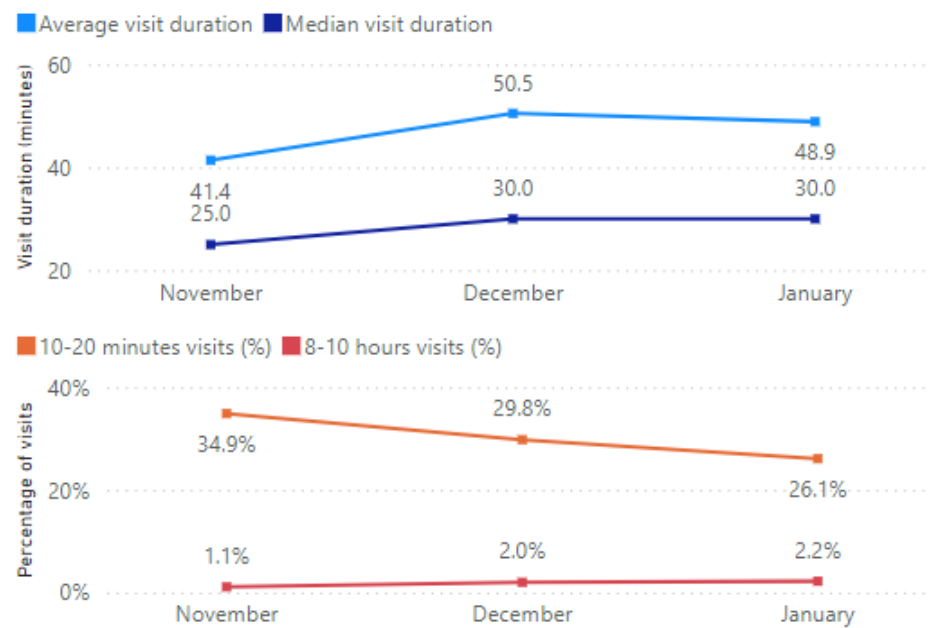


Figure 10. KPI evolution.

The dashboard prototyping and preliminary analysis have shown to PBL Care the merit and capabilities of our proposed digitisation approach in monitoring KPIs of SMEs through interactive dashboards with the aim of improving the quality of services provided to patients.

5.1.4. Findings—Improving Provision of Care through Data Reporting

Following the creation of the dashboard prototype, PBL Care decided to use it during monthly meetings with care staff to show progress and set actions. Based on the collected data, dashboard visualisations can highlight problems in an organisation and lead to the modification and implementation of new processes. These steps can then be repeated and, as the number of iterations grows, it is likely to observe improvements in the service delivery based on continual improvement processes.

Conscious of the importance of collecting data digitally, PBL Care started using an accredited software provider from February 2019, including a mobile application for care staff and a web application for the PBL Care management team. In March 2019, the Care Quality Commission conducted a new inspection of PBL Care which resulted in a rating of “Good”, improved from the December 2017 rating “Requiring improvement”.

We have learnt that a digital solution can help monitor the provision of care. However, it must not be the only source of information for driving decisions. Underlying factors can be at play behind the scenes and not be reflected on the dashboard. The digital tool can help in generating hypotheses and highlighting patterns that need to be raised in conversations between management staff, care staff and service users. For example, the management team must not jump to conclusions and blame care staff if the dashboard shows that a couple of visits lasted less than 10 min. There could be a rational explanation, for example a patient could have told the carer he or she did not require care on that day. The identification of KPIs and interpretation of data must not be based only on performance, it must also reflect the provision of good care and satisfaction of individuals, both service users and care staff.

5.2. Parameter Optimisation for Additive Manufacturing Using Supervised Machine Learning

Additive manufacturing (AM) is the process of fabricating components by adding layer upon layer of materials with the aid of digital 3D design [34]. AM has recently gained increasing research attention because of its advantages in comparison with traditional subtractive manufacturing [35]. Although the use of machine learning for additive man-

ufacturing is still in its infancy, several machine-learning algorithms have been applied in AM tasks including parameter optimisation and fault detection [24,34]. Related to this, Meng et al. [24] review the latest applications of different machine learning algorithms in additive manufacturing. The study matches various ML methods to corresponding AM applications including parameter optimisation and anomaly detection.

This case study is about parameter optimisation in 3D printing for a company called HiETA Technologies. It attempts to investigate a model for establishing the relationship between specified process inputs and defect indicators in produced components. HiETA Technologies is a research and product development company founded in 2011. The SME specialises in the use of additive manufacturing (metal 3D printing) for thermal management and light-weighting solutions. They offer an end-to-end metal 3D printing service in additive manufacturing and engage in development projects for a range of energy systems including fuel cells, turbine machinery, nuclear, concentrated solar power, and other heat and power generation systems for automotive and aerospace applications. HiETA's unique technical capabilities and technologies include the development of heat-transfer surfaces with increased high heat transfer. The company strives to create methods and processes that dramatically reduce product size, and increase cycle efficiencies and product life.

5.2.1. Problem Statement

HiETA Technologies currently uses a series of trial and error testing (printing component samples repeatedly and testing their quality and defects) to optimise process parameters and reach optimal input values for the production of the final components. This brute force method, commonly used in the wider 3D printing business sector, not only costs a significant amount of time causing major delays, but also incurs countless failures and mistakes before yielding the right parameters that work for the given production. To address this issue, the company wants to investigate the possibility of using machine learning in their processes to reduce errors and the high time and material costs associated with the use of their existing trial and error method. On that initiative, HiETA joined the BDC project in a research collaboration and provided a pilot dataset to investigate this initiative.

5.2.2. Initial Data Preparation and Exploration

HiETA's primary objective of applying machine learning in its additive manufacturing process is to automatically identify the optimal parameters for building components with few or no trials. In particular, the task of this first stage is to develop a bespoke predictive model to be used in the material characterisation of new product development. This will in the long term optimise the process parameters for a given geometry–material–machine combination, thus improving the SME's understanding of parameter interactions and cause–effect relationships in the AM process. Such a predictive model will also establish a relationship formula between the input and target parameters enabling HiETA to run the minimum possible experiments to obtain the right parameter values that work. This will also focus production time and resources into a small pool of experimental trials.

HiETA provided the BDC project with sample data of six parameters for this pilot investigation. Table 2 shows an anonymised statistical summary of the sample parameters' data and their modelling roles. You will note that four of the data parameters, namely laser power (LP), point distance (PD), hatch option (HO) and exposure time (ET), are predictors whereas border length densities (S1BLD, S2BLD and S3BLD) and bulk length densities (S1BULD, S2BULD and S3BULD) for samples 1, 2 and 3 are response variables. The table shows that all four input parameters are normally distributed as shown in their mean–median equality and their skewness. This suggests that there is no need for performing any variable transformation prior to training a predictive model. The variables with roles indicated by output (excluded) are not included in the modelling exercise as these were not a priority for the SME at the time of this research collaboration. Thus, the four predictors (LP, PD, HO and ET) and the two SME prioritised target variables (S1BLD and S2BLD) are

extracted from the sample data to investigate a predictive model for minimising the defect densities in samples 1 and 2.

Table 2. Summary statistics of the sample data parameters and roles (rounded to 3 S.F).

Parameter	Mean	Std Deviation	Median	Skewness	Role
ET	0.645	0.800	0.645	0.000	Input (predictor)
HO	0.387	0.533	0.387	0.000	Input (predictor)
LP	1.806	1.600	1.806	0.000	Input (predictor)
PD	1.161	1.067	1.161	0.000	Input (predictor)
ABULD	0.752	0.583	0.851	0.638	Output (excluded)
ALD	1.248	1.339	1.167	1.163	Output (excluded)
S1BLD	1.359	1.698	1.185	1.237	Output (response)
S1BULD	0.429	0.430	0.394	0.619	Output (excluded)
S2BLD	1.313	1.362	1.184	1.175	Output (response)
S2BULD	1.149	0.844	1.436	0.540	Output (excluded)
S3BLD	1.073	1.094	1.045	1.194	Output (excluded)
S3BULD	0.677	0.651	0.738	1.435	Output (excluded)

Figure 11 summarises the implementation workflow we used to investigate the development of the predictive model. Data exploration was the first step of the analytic workflow, which examined data characteristics including the importance and correlations of the input variables to the output parameters. We have also looked at the importance of the four process inputs to the two target variables. Figure 12 (cf. Table 3) shows the linear correlations between the four process input parameters in the sample dataset and the target variables. It is evident that all four input parameters are better predictors for sample 1 border length density (S1BLD) compared to the ones from sample 2 (S2BLD). It is also clear that the HO and PD process inputs are more important than LD and ET in predicting both defect densities, S1BLD and S2BLD. Overall, the low linear correlation r-values given in Table 3 suggest that nonlinear machine learning models may better predict the product defects than their linear counterparts, as we will see later (Results and Discussion section).

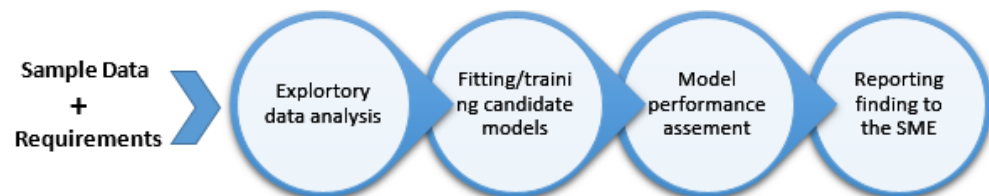


Figure 11. Parameter modelling workflow.

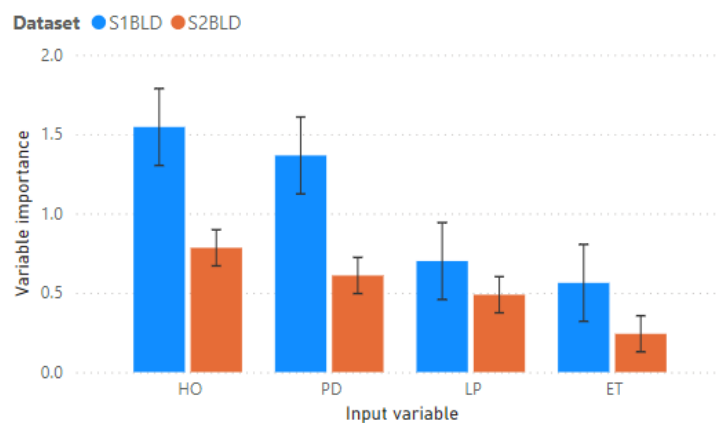


Figure 12. Importance of process inputs for the prediction of output parameters (targets).

Table 3. Input variable correlations with response variables.

Input Variables	Target Variables	
	S1BLD	S2BLD
Exposure Time (ET)	0.13798	0.12041
Laser Power (LP)	0.09345	0.11580
Point Distance (PD)	−0.19764	−0.07931
Hatch Option (HO)	−0.41878	−0.39791

5.2.3. Modelling Process Parameters Using Regression Algorithms

The target variables of interest in this predictive modelling task (S1BLD and S2BLD) are continuous-valued. This makes regression approaches the most appropriate predictive modelling algorithms to be used. To this end, we have applied six different supervised algorithms to the sample data, namely the linear, polynomial, decision tree, random forest, SVM and Multilayer Perceptron regression methods. The following two sections respectively present a brief overview of the applied regression methods and their application to the sample process input–output parameter data.

Applied Machine Learning Algorithms

First, linear regression is one of the most widely used supervised machine learning algorithms, and models the relationship between a numeric target and input variable(s) by fitting linear Equation (1) to observed data. Such an equation is then used to make a prediction by computing a weighted sum of the input features and an intercept term (α —the value of the target variable \hat{y} when input values are not present, i.e. all $x_i = 0$).

$$\hat{Y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad (1)$$

where \hat{Y} is the dependent (predicted) variable and β_i is the coefficient of the i^{th} model input (independent) parameter (x_i).

Second, one primary shortfall of linear regression models is the assumption that a linear relationship exists between the predicted and input variable(s). This does not necessarily always hold, which is why we have considered experimenting with polynomial and Support Vector Regression (SVR) methods to model the process input–output relationship in these data. Polynomial regression works like linear regression except that it adds powers to each predictor as a new feature, for example, if we have a single input in the model, $\hat{y} = \alpha + \beta_1 x_1$, a second-degree polynomial term changes it to $\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_1^2$. On the other hand, SVR is a derivative of the popular Support Vector Machine (SVM), one of the most powerful and versatile supervised learning algorithms for both linear and nonlinear regression problems [36]. SVR uses the same principles as the SVM kernels and hyperparameters such as the cost and error functions. In our work, we employed Radial Basis Function (RBF), a common kernel widely used to address nonlinearity problems in datasets. The RBF-based SVR algorithm transforms the data by creating new features from the nonlinear data and estimates the target values as per Equation (2).

$$\hat{Y} = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (2)$$

where α_i are the dual coefficients, $K(\exp(\gamma||x - x_i||^2))$ is the RBF kernel function and b is the intercept.

Third, Neural Networks (NNs) are also a set of powerful supervised learning algorithms used for regression modelling. In the context of this study, we used the Multilayer Perceptron (MLP), a simple but efficient technique consisting of an input layer with n neurons (inputs), a hidden layer and an output layer (output). When used for predict-

ing continuous-valued data as in our case, the MLP model approximates the functional relationship between the input and response variables as per expression (3).

$$\hat{Y} = f(w_0 + \sum_{j=1}^m w_j f(w_{0j} + \sum_{i=1}^n w_{ij} x_i)) \quad (3)$$

In Equation (3), w_0 is the intercept of the output neuron and w_j is the weight from the j^{th} hidden neuron to the output layer.

Finally, tree-based models including decision trees (DTs) and random forest (RF) are another family of popular learning techniques for regression problems. In addition to their insensitivity to anomalies including missing data and outliers, these methods also capture nonlinear data relations. Regression DTs partition the predictor variables into clusters (branches) by optimising an objective function such as the Mean Error Square (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{Y})^2 \quad (4)$$

where n is the training data at a given node, y_i is the actual target value and $\hat{Y}(\frac{1}{n} \sum_{i=1}^n y_i)$ is the mean predicted value in the node. The random forest algorithm is a combination of multiple decision trees and is known to have a better generalisation performance compared to models built with single decision trees [36]. To arrive at a prediction from multiple DTs, RF for regression averages the outputs of the different constituent decision trees.

Results and Discussion

The learning algorithms discussed in the Applied Machine Learning Algorithms section require a sufficient amount of training data to produce good and reliable predictive models. However, one challenge facing HiETA, and perhaps the wider AM research and product development companies, is the lack or insufficiency of past fabrication data, which led some researchers to suggest the use of simulated data for ML-based parameter optimisation [24]. Insufficient training data are a common challenge in machine learning, particularly for domains starting to adopt AI or in which the process of creating the training data is very expensive and time consuming such as additive manufacturing [24,37]. To mitigate that, we have used the SMOGN (Synthetic Minority Over-Sampling Technique for Regression with Gaussian Noise) algorithm [38] to oversample the data into various sizes and up to 2000 instances as the original data was just less than 100 observations. It is widely accepted in AI/ML that predictive accuracy improves with increased training data [39]. Resampling limited training data and synthesising new data instances are common practices in predictive machine learning with the objective of increasing small-sized datasets.

To build the product defect predictive model, we have applied six regression algorithms (Applied Machine Learning Algorithms section) to the oversampled data of up to 2000 instances. Tables 4 and 5 summarise the Root Mean Square Error (RMSE) results of the models built with these algorithms and the best oversampling rates, 500 and 1000. The RMSE is a regression performance measure used to evaluate the differences between actual and model predicted parameter values. It is computed as per Equation (5), where n , y_i , \hat{y}_i represent the sample size, and actual and predicted values, respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

The results (Tables 4 and 5) show that decision trees, random forest and polynomial models perform better for this sample data compared to other applied regression models. Polynomial models are known for their high computational complexity with increasing

data variables and degrees. Overall, it is clear from this set of results, that decision trees produce the lowest prediction error consistently across all models and for all oversampling rates (this includes results achieved with other oversampled data sizes above 1000 and up to 2000 instances). The fact that linear regression has the lowest accuracy among all models suggests that the relationship between manufacturing defects and input parameters is nonlinear. This corroborates the findings in the initial data exploration, e.g., the correlation coefficients between the predictors and target parameters as in Table 3.

Table 4. Defect prediction performance of different regression algorithms based on oversampled data of size 500 instances: all figures are rounded to 3 S.F.

Algorithm	S1BLD	S2BLD	S1BULD	S2BULD
Baseline	1.475	1.046	0.474	0.865
Linear	1.32	0.926	0.384	0.74
Polynomial	0.485	0.259	0.117	0.231
Decision Tree	0.410	0.25	0.090	0.126
Random Forest	0.461	0.289	0.104	0.161
SVM (rbf)	0.777	0.423	0.139	0.303
Neural Network	0.598	0.314	0.140	0.251

Table 5. Defect prediction performance of different regression algorithms based on oversampled data of size 1000 instances: all figures are rounded to 3 S.F.

Algorithm	S1BLD	S2BLD	S1BULD	S2BULD
Baseline	1.718	1.188	0.513	0.954
Linear	1.282	1.025	0.380	0.786
Polynomial	0.433	0.293	0.144	0.286
Decision Tree	0.379	0.257	0.092	0.114
Random Forest	0.404	0.284	0.106	0.140
SVM (rbf)	0.545	0.363	0.148	0.335
Neural Network	0.466	0.753	0.165	0.285

In Table 6, we show the average performance of all algorithms across varying oversampled data sizes including the original one (the size of the original data could not be disclosed due to the SME’s business confidentiality). Figure 13 visually illustrates the data size effect on prediction performance as presented in Table 6. Empirically speaking, the oversampled data in the range between 500 and 1000 produces the best overall average accuracy across models. In other words, the average prediction errors produced with oversampled datasets 1500 and 2000 for nearly all models are higher than those achieved with lower level oversampling, e.g., the 500 and 1000 observations. This may be due to the fact that any noise in the data becomes significantly amplified at the higher oversampling rates.

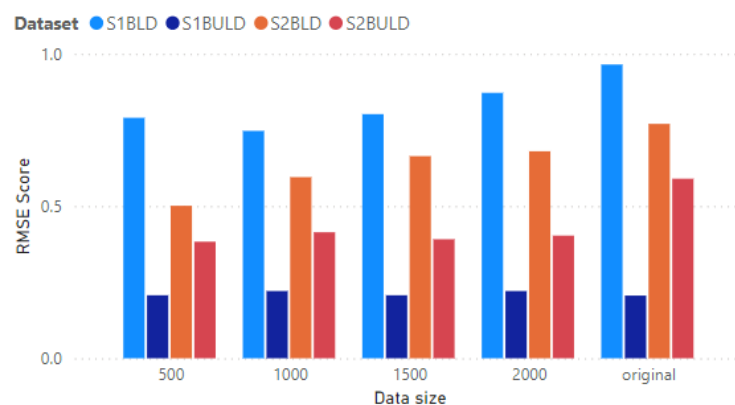


Figure 13. Oversampled data size effect on model performance.

Table 6. Average dataset performance over all regression algorithms.

Data Size	S1BLD	S2BLD	S1BULD	S2BULD
original	0.965	0.770	0.206	0.590
500	0.790	0.501	0.207	0.383
1000	0.747	0.595	0.221	0.414
1500	0.802	0.664	0.207	0.391
2000	0.872	0.680	0.221	0.403

To investigate any possible further improvement that might be achieved with the best identified model, we have tuned the decision tree depth hyperparameter. For illustration purposes only, Table 7 and Figure 14 show the model errors (hence accuracy) for the S1BLD target variable for different oversampled training data sizes while varying decision tree depths. The results show that model accuracy improves with increasing decision tree depth, although the errors become stable from a maximum depth of 21. In addition, the oversampled data of size 1500 instances consistently achieves the lowest error with various models.

Table 7. Results of decision tree models for S1BLD target with varying maximum depth and oversampled data sizes.

Tree Depth	500	1000	1500	2000
3	0.779	0.711	0.768	0.857
6	0.512	0.455	0.442	0.500
9	0.446	0.400	0.378	0.419
12	0.420	0.383	0.355	0.387
15	0.415	0.379	0.346	0.377
18	0.411	0.379	0.343	0.373
21	0.410	0.379	0.343	0.372
24	0.410	0.379	0.343	0.371
27	0.410	0.379	0.343	0.371
30	0.410	0.379	0.343	0.371

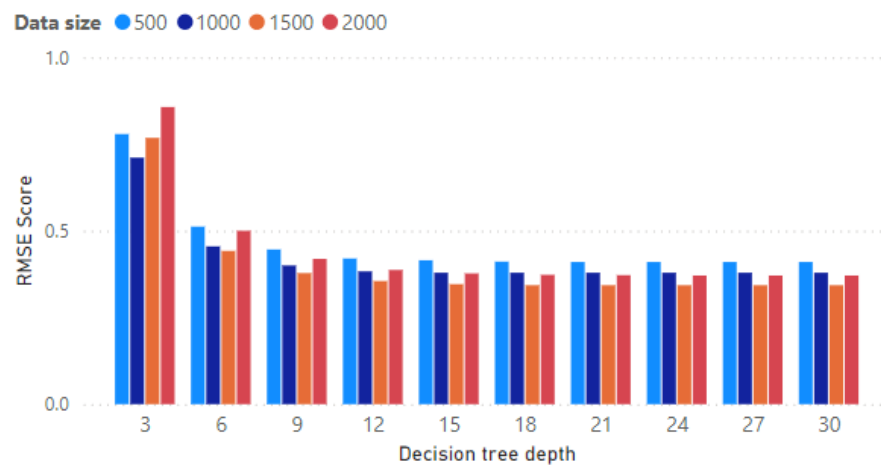


Figure 14. Effect of tuning decision tree depth on model performance.

Based on our analysis, there are several reasons for the weakness of the tested models with the given sample data size ($n < 100$) being the primary one. The analytic sample data were also found to be noisy (confirmed by the SME), which also contributed to the high model errors. Statistically speaking, it is widely accepted in the literature to expect a high error margin when using a small sample dataset with a statistical model, particularly with multiple predictors [40]. Overall, with less noisy data in place, we think this will be a good

step in paving the way towards a fully fledged parameter optimisation predictive tool to be used within HiETA's AD process and new product development. It was also encouraging to learn that another independent (commercial) investigation of the sample data concluded similar findings based on the SME feedback to us.

6. Summary and Conclusions

Digitalisation in SMEs has never been more needed than in the post-COVID-19 era as most businesses adopted online and hybrid operations with many of their employees working from home. In fact, earlier studies in the pandemic have found that up to 70% of SMEs stepped up digital technology use during the COVID-19 era [14]. However, considering constraints such as the lack of sufficient finance to invest in IT infrastructure and to hire the right skilled experts, SMEs are still far from being in full swing as part of the digital revolution.

Evidently, there is a need to raise awareness among SME owners, managers and entrepreneurs about the advantages and challenges digitalisation could bring to their business, and how different subfields of data science could apply to different industries, business functions and business models. Decision makers have to train in order to rethink their business processes and to reconfigure tasks and organisational structures. More staff would also require upskilling in order to consider and guide analytical outputs, and take a data-driven approach to solve problems leading to more informed decisions that benefit the business. The range of challenges identified as a result of our analysis includes the following:

- Need to support SMEs in building a culture of data, from collection, to management, to protection and processing, in order to ensure that the digitisation transition takes place with the least risk to SMEs.
- Raise awareness of the benefits of data science and analytics to the business.
- Upskill SME managers and employees, ensuring an involved approach for redesigning business processes and training required to run applications, and analysing results.
- Consider mechanisms to bridge the financing gap until the data science solution can deliver its full potentials.
- Enable SMEs to gradually increase their capacity before being eventually able to develop their own data science solutions.
- Provide an analysis of the sectoral impact of data science on SMEs' business activities, with specific business use cases, and inform pertinent investors.
- Better understand the role that business associations, chambers of commerce, academia, national and local governments, international organisations and other SMEs could play to progress on these different dimensions, and support with knowledge sharing and open data availability.

Our analysis clearly suggests that most SMEs collect and store some sort of business data but require skills to analyse and produce useful insights for data-driven decision making. If SMEs are empowered with the right skills and/or supported financially for this purpose, they can make full utilisation of data to help them, hence driving the growth of the entire economy.

Author Contributions: Conceptualization: M.M. and A.-R.H.T.; methodology: M.M.; data analysis: M.M., X.S. and K.V.; writing: A.-R.H.T., M.M., X.S., K.V. and D.H.; visualization: X.S., K.V. and M.M.; supervision: A.-R.H.T. and M.M.; funding acquisition: A.-R.H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the European Union under the European Regional Development Fund (The Big Data Corridor project—Project no. 12R16P00220) and match-funded by six Project Partners—Birmingham City Council, Aston University, Birmingham City University, You Smart Thing, Innovation Birmingham and West Midlands Combined Authority. The funding source had no role in the design, execution, interpretation or writing of the work.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from third party and are available from the authors with the permission of the third party.

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

References

1. UK Small Business Statistics. Available online: <https://www.merchantsavvy.co.uk/uk-sme-data-stats-charts/> (accessed on 10 June 2024).
2. Hutton, G. *Business Statistics*; House of Commons Library: London, UK, 2024. Available online: <https://researchbriefings.files.parliament.uk/documents/SN06152/SN06152.pdf> (accessed on 16 June 2024).
3. Lam, S.K.; Sleep, S.; Hennig-Thurau, T.; Sridhar, S.; Saboo, A.R. Leveraging frontline employees' small data and firm-level big data in frontline management: An absorptive capacity perspective. *J. Serv. Res.* **2017**, *20*, 12–28. [\[CrossRef\]](#)
4. Mohamed, M.; Weber, P. Trends of digitalization and adoption of big data & analytics among UK SMEs: Analysis and lessons drawn from a case study of 53 SMEs. In Proceedings of the 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 15–17 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
5. Ragazou, K.; Passas, I.; Garefalakis, A.; Galariotis, E.; Zopounidis, C. Big data analytics applications in information management driving operational efficiencies and decision-making: Mapping the field of knowledge with bibliometric analysis using R. *Big Data Cogn. Comput.* **2023**, *7*, 13. [\[CrossRef\]](#)
6. Alvarez, I.; Zamanillo, I.; Cilleruelo, E. Have information technologies evolved towards accommodation of knowledge management needs in Basque SMEs? *Technol. Soc.* **2016**, *46*, 126–131. [\[CrossRef\]](#)
7. Lee, J.W. Analysis of technology-related innovation characteristics affecting the survival period of SMEs: Focused on the manufacturing industry of Korea. *Technol. Soc.* **2021**, *67*, 101742. [\[CrossRef\]](#)
8. Nasrollahi, M.; Ramezani, J.; Sadraei, M. The impact of big data adoption on SMEs' performance. *Big Data Cogn. Comput.* **2021**, *5*, 68. [\[CrossRef\]](#)
9. Wang, S.; Wang, H. Big data for small and medium-sized enterprises (SME): A knowledge management model. *J. Knowl. Manag.* **2020**, *24*, 881–897. [\[CrossRef\]](#)
10. Gupta, S. *Driving Digital Strategy: A Guide to Reimagining Your Business*; Harvard Business Press: Boston, MA, USA, 2018.
11. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [\[CrossRef\]](#)
12. Marcinkowski, B.; Gawin, B. Data-driven business model development—insights from the facility management industry. *J. Facil. Manag.* **2020**, *19*, 129–149. [\[CrossRef\]](#)
13. Khayer, A.; Talukder, M.S.; Bao, Y.; Hossain, M.N. Cloud computing adoption and its impact on SMEs' performance for cloud supported operations: A dual-stage analytical approach. *Technol. Soc.* **2020**, *60*, 101225. [\[CrossRef\]](#)
14. OECD. The Digital Transformation of SMEs. In *OECD Studies on SMEs and Entrepreneurship*; OECD Publishing: Paris, France, 2021. [\[CrossRef\]](#)
15. Willetts, M.; Atkins, A.S.; Stanier, C. Barriers to SMEs adoption of big data analytics for competitive advantage. In Proceedings of the 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21–23 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
16. Coleman, S.; Göb, R.; Manco, G.; Pievatolo, A.; Tort-Martorell, X.; Reis, M.S. How can SMEs benefit from big data? Challenges and a path forward. *Qual. Reliab. Eng. Int.* **2016**, *32*, 2151–2164. [\[CrossRef\]](#)
17. Bhardwaj, S. Data Analytics in Small and Medium Enterprises (SME): A Systematic Review and Future Research Directions. *Inf. Resour. Manag. J.* **2022**, *35*, 1–18. [\[CrossRef\]](#)
18. Bakhshi, H.; Bravo-Biosca, A.; Mateos-Garcia, J. *The Analytical Firm: Estimating the Effect of Data and Online Analytics on Firm Performance*; Nesta Working Paper No. 14/05; Nesta: London, UK, 2014.
19. Ghasemaghaei, M. Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *Int. J. Inf. Manag.* **2019**, *57*, 102055. [\[CrossRef\]](#)
20. UK Government. 2500 New places on Artificial Intelligence and Data Science Conversion Courses. Available online: <https://www.gov.uk/government/news/2500-new-places-on-artificial-intelligence-and-data-science-conversion-courses-now-open-to-applicants> (accessed on 10 June 2024).
21. Schönberger, M. Artificial Intelligence for Small and Medium-sized Enterprises: Identifying Key Applications and Challenges. *J. Bus. Manag.* **2023**, *21*, 89–112. [\[CrossRef\]](#)
22. Griesch, L.; Rittelmeyer, J.; Sandkuhl, K. Towards AI as a Service for Small and Medium-Sized Enterprises (SME). In Proceedings of the IFIP Working Conference on the Practice of Enterprise Modeling, Vienna, Austria, 28 November–1 December 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 37–53.
23. Qalati, S.A.; Yuan, L.W.; Khan, M.A.S.; Anwar, F. A mediated model on the adoption of social media and SMEs' performance in developing countries. *Technol. Soc.* **2021**, *64*, 101513. [\[CrossRef\]](#)
24. Meng, L.; McWilliams, B.; Jarosinski, W.; Park, H.Y.; Jung, Y.G.; Lee, J.; Zhang, J. Machine learning in additive manufacturing: A review. *JOM* **2020**, *72*, 2363–2377. [\[CrossRef\]](#)

25. Mancini, J. *Data Portability, Interoperability and Digital Platform Competition: OECD Background Paper*; OECD: Paris, France, 2021.
26. Banerjee, A.; Bandyopadhyay, T.; Acharya, P. Data analytics: Hyped up aspirations or true potential? *Vikalpa* **2013**, *38*, 1–12. [[CrossRef](#)]
27. Soroka, A.; Liu, Y.; Han, L.; Haleem, M.S. Big data driven customer insights for SMEs in redistributed manufacturing. *Procedia CIRP* **2017**, *63*, 692–697. [[CrossRef](#)]
28. PBL Care. Available online: <https://pblcare.com/> (accessed on 10 June 2024).
29. Care Quality Commission. Available online: <https://www.cqc.org.uk/> (accessed on 10 June 2024).
30. Atasoy, H.; Greenwood, B.N.; McCullough, J.S. The digitization of patient care: A review of the effects of electronic health records on health care quality and utilization. *Annu. Rev. Public Health* **2019**, *40*, 487–500. [[CrossRef](#)] [[PubMed](#)]
31. Mihailescu, M.; Mihailescu, D. The emergence of digitalisation in the context of health care. In Proceedings of the 51st Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 3–6 January 2018. [[CrossRef](#)]
32. Soikkeli, J.; Pulkkinen, M.; Ruohonen, T. Evaluating the Value of Enterprise Resource Planning in Home Care Services. In Proceedings of the European Conference on Information Systems Management, Utrecht, The Netherlands, 6–8 June 2013; Academic Conferences International Limited: Reading, UK, 2013; p. 249.
33. Home Care: Delivering Personal Care and Practical Support to Older People Living in Their Own Homes. Available online: <https://www.nice.org.uk/guidance/ng21/resources/home-care-delivering-personal-care-and-practical-support-to-older-people-living-in-their-own-homes-pdf-1837326858181> (accessed on 10 June 2024).
34. Delli, U.; Chang, S. Automated process monitoring in 3D printing using supervised machine learning. *Procedia Manuf.* **2018**, *26*, 865–870. [[CrossRef](#)]
35. Qi, X.; Chen, G.; Li, Y.; Cheng, X.; Li, C. Applying neural-network-based machine learning to additive manufacturing: Current applications, challenges, and future perspectives. *Engineering* **2019**, *5*, 721–729. [[CrossRef](#)]
36. Raschka, S.; Liu, Y.H.; Mirjalili, V.; Dzhulgakov, D. *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*; Packt Publishing Ltd.: Birmingham, UK, 2022.
37. Lateh, M.A.; Muda, A.K.; Yusof, Z.I.M.; Muda, N.A.; Azmi, M.S. Handling a small dataset problem in prediction model by employ artificial data generation approach: A review. *Proc. J. Phys. Conf. Ser.* **2017**, *892*, 012016. [[CrossRef](#)]
38. Branco, P.; Torgo, L.; Ribeiro, R.P. SMOGN: A pre-processing approach for imbalanced regression. In Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications, Skopje, Macedonia, 22 September 2017; PMLR: London, UK, 2017; pp. 36–50.
39. Halevy, A.; Norvig, P.; Pereira, F. The unreasonable effectiveness of data. *IEEE Intell. Syst.* **2009**, *24*, 8–12. [[CrossRef](#)]
40. Kroll, C.N.; Song, P. Impact of multicollinearity on small sample hydrologic regression models. *Water Resour. Res.* **2013**, *49*, 3756–3769. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.