Portfolio management: The holistic data lifecycle

Portfolio Management: The Holistic Data Lifecycle

John McAvoy
University College Cork, Ireland

Conor Murphy
University College Cork, Ireland

Laila Mushtaq
University College Cork, Ireland

James O’Donnell
University College Cork, Ireland

Attracta Brennan
University of Galway, Ireland

Mary Dempsey
University of Galway, Ireland

Gaye Kiely
University College Cork, Ireland

ABSTRACT

Machine learning provides many benefits to Portfolio Managers in analysing data and has the potential to provide much more. A concern with the approach to Machine Learning in Portfolio Management is that is caught between two domains: finance and information systems. In reality, to ensure its success, having these two separate and distinct domains are problematic. What is required is a holistic view, facilitating discussions, with data being the unifying concept and the one that is key to success. The data value map is a lens that allows all involved, in the use or adoption of Machine Learning in Portfolio Management, to form a shared understanding of the lifecycle of the data involved. Rather than being seen as a financial concept or a technical concept, this view of the data lifecycle provides a platform for all involved to determine what is required, and to identify and deal with any potential pitfalls along the way. A holistic view, and shared understanding, are required for the success of Machine Learning in Portfolio Management. Research on the intersection between Machine Learning and Portfolio Management is currently lacking. A focus on the different parts of the data lifecycle provides an opportunity for further research.

Subjects: Finance, Information Systems

INTRODUCTION

There has been a large increase in the use of Machine Learning (ML) techniques and growing interest in their applications in finance, which have been applied primarily in Portfolio Management (Perrin, 2019). There are many articles showing the benefits of, and how to use, ML as a technology for investing (cf. Das et al., 2021; Li, Simon and Turkington, 2022). Although benefits can be seen for Machine Learning in different areas such as asset allocation portfolios (Konstantinov, Chorus and Rebmann, 2020), detecting economic regimes (Mulvey and Liu, 2016), and hedge funds (Stentoft and Wang, 2020), there has been a lot of failure in its use (López de Prado, 2018). With the availability and understanding of data being at the forefront of Portfolio Management, LaValle (2011) finds that having a clear understanding of the importance of data projects and how they impact business value often gets overlooked by organizations. At the core of these projects is the underlying data, and organisations need a way of understanding and representing that data. Tools such as the Data Value Map can help organizations visualize where their data is impacting their business value. The Data Value Map is a visualization tool that helps to build shared understanding around data initiatives. The Data Value Map can help remedy the disconnect between factors such as the shortage of communal understanding and misalignment amongst stakeholders of the data, resulting in an organisations' inability to leverage their data's full business value (Nagle and Sammon, 2017).

This paper looks at the criteria necessary for the implementation of Machine Learning in Portfolio Management, and analyses how these criteria impact business value.

RESEARCH METHODOLOGY

This paper uses the guidelines and instructions demonstrated by Webster and Watson (2002) to conduct an extensive literature review on the intersection of ML and portfolio management. A two-phase process was followed to establish a grouping of articles that form a literary base for analysis and investigation. In the first step, a review of previous literature was conducted to acquire a general understanding of the current field of study. In the second step, these articles were thoroughly examined to refine the literature base only to papers which contain a high degree of relevance to the research question. Literature was read and examined separately by the authors and then combined to validate.

According to Vom Brocke (2009), the first vital step in analysing pre-existing literature is to review leading journals and journal databases. The searches were limited to papers published between 2000 and 2020 to take into account the relative recency of research in this area. As this paper covers topics outside of just portfolio management, other disciplinary journals in the areas of information systems, economics, and technology were used.

To build the initial literature database, extensive literature searches, through a cluster of dominant journals in the IS field, collectively known as the "basket of eight" was performed. The 'basket of eight' comprises the eight leading IS academic Journals which hold the most prestige and attract the highest level of academic research in this field. The College of Senior Scholars has identified the journals in the "basket" as the top journals in the IS field (AIS, 2011).

As this paper covers topics outside IS, other disciplinary journals in the areas of finance, economics, and technology were used. These consist of the Quarterly Journal of Economics, the European Journal of Operational Research, the Journal of Financial Economics, and the Journal of Machine Learning Research. The Journal of Financial Economics was chosen because it places high emphasis in the areas of financial institutions and corporate finance (Elsevier, 2020). As this research is looking at the criteria necessary for Machine Learning to be implement in the Portfolio Management industry, it was felt that this journal would give a better insight into the finance industry and how ML operations already work in the finance sector. The European Journal of Operational Research publishes high quality papers in the area of decision making (Elsevier, 2020) and was chosen to gain an understanding of the decision making process of the Portfolio Management industry when implementing ML. The Journal of Machine Learning Research was chosen because it specialises in Machine Learning. Similar to The Journal of Financial Economics and The Quarterly Journal of Economics was chosen because it provides a wide range of articles on ML in Finance. The journal has a 2018 impact factor of 11.775, ranking it 1st out of 363 journals in the category of "Economics" (Oxford Academic, 2020).

To find key studies from the wide range of articles identified, and to determine what articles to include and exclude, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses or (PRISMA) technique developed by Moher et al (2009) was employed, with our approach shown in Figure 1. Studies were conducted using a titles/abstracts/keywords search using the following keywords and their synonyms, to ensure that the research question was reflected in the literature search the following terms were used: "Machine Learning", "Adoption", "Criteria", "Portfolio Management" & "Investment Management".

The "Check for duplicates" feature in Mendeley Bibliography Reference Management Software returned 6,408 articles which were then considered suitable candidates for additional screening. Following this, steps described by Webster and Watson (2002) were followed to filter appropriate content out for further investigation. White papers from companies were included to give a perspective from industry,

These papers were then further subjected to an examination where their abstract and introduction were read. When in doubt, a full-text review was performed to determine a paper's suitability, accepting an article only if it identifies concepts relevant to the research question. This further reduced the set from 505 papers to 59. These figures show that this is a research area where Machine Learning is mentioned in a lot of financial services' research but has limited research that concentrates solely on the topic.

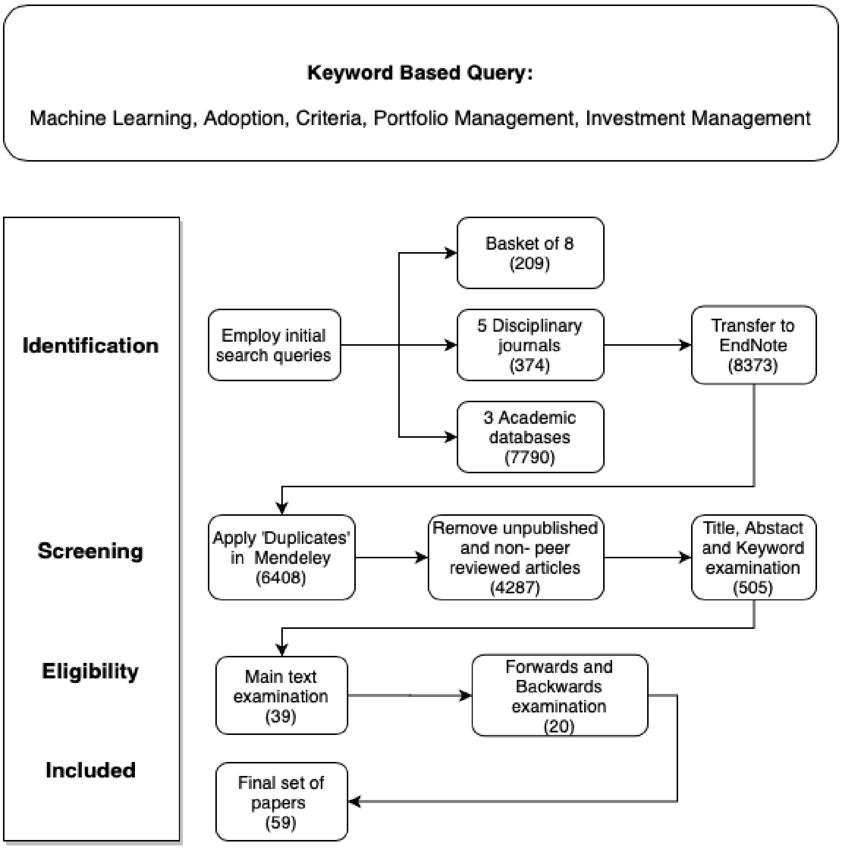


Figure 1. Literature Review Methodology

Concept-Centric Matrix

Following Webster and Watson (2002) instructions, a concept centric matrix (Table 1) was formulated, as a means of representing the analysis of the literature. The set of articles that qualified for in-depth analysis were examined and reviewed in full and corresponding key concepts were identified (Creed et al., 2021).

Concepts	Unit of Analysis	Number of Citations	Citations
Cost	Infrastructure to implement	[6]	(Joshi and Wheelock, 2018), (FSB, 2017), (Goldman, 1999), (CFA, 2019), & (Vanson Bourne, 2015)
	Financial Restriction	[6]	(Michalkova, 2018), (Deloitte, 2018), (Savignac, 2008), (Cooper et al., 2001). & (Pwc, 2018)
People	People	[7]	(Walter and Zimmermann, 2016), (Discenza and Forman, 2007), (LinkedIn, 2017), (Lorica, 2018), & (FSB, 2017)
Data	Availability of Quality Data	[12]	(Cooper et al., 2002), (Kock and Gemünden, 2016), (Martinsuo and Kilen, 2014), (Martinsuo and Lehtonen, 2007), (Jonas et al., 2013), (Voss and Kock, 2013), (Stork, 2000), (Pwc, 2018), (Corrales, 2018), (DuBois, 2004), (Mohajan, 2017), & (Bean, 2018)
	Access to Data	[4]	(Ryll et al, 2020), (Martens, 2018), (FSB, 2017), & (Ali et al, 2016)
Process	Managerial Vision	[3]	(Van Den Steen, 2005), & (CFA, 2019)
	ML Malpractice	[13]	(Ait Gougam et al,, 2007), (Harvey and Liu, 2015), (Sharpe, 1994), (Bailey et al., 2014), (Ioannidis, 2005), (Hawkins, 2004), (Cvitanic et al, 2007), (Cohen and Jensen, 1997), (Papernot et al., 2016), & (Beel and Brunel, 2019)
	Confidential Business Information	[6]	(Almeling, 2012), (Dobrusin, 2012), (Sonnenburg et al., 2007), (Robertson, 2015), & (Vanson Bourne, 2017)
Black Box	Black Box	[5]	(Krause, 2016), (Gillespie, 2014), (Van Liebergen, 2017), (Cabitza, 2017), & (Deloitte, 2018)

Table 1. Core factors in research into ML and portfolio management

These factors represent the core themes in existing research, The next step was to analyse them from the perspective of how this can drive value for portfolio management. The data value map was used as a tool to analyse and catagorise the data.

Data Value Map

While being integral to unlocking the value of their data, ML can also provide the Portfolio Management industry with greater accuracy, speed, and dependability and can further generate larger productivity and profitability than less innovative counterparts (Hewitt-Dundas, 2006). To facilitate a discussion around this, this study employs the Data Value Map created by Nagle and Sammon (2017), as seen below in figure 2. The Data Value map is introduced as an artefact designed to remedy the disconnect between factors, resulting in an organisations' inability to leverage the full value of its data.

The Data Value Map (DVM) has an emphasis on the value that is born from data initiatives. In every instance, value is assumed but, in reality, is not always applied, according to Nagle and Sammon (2017).

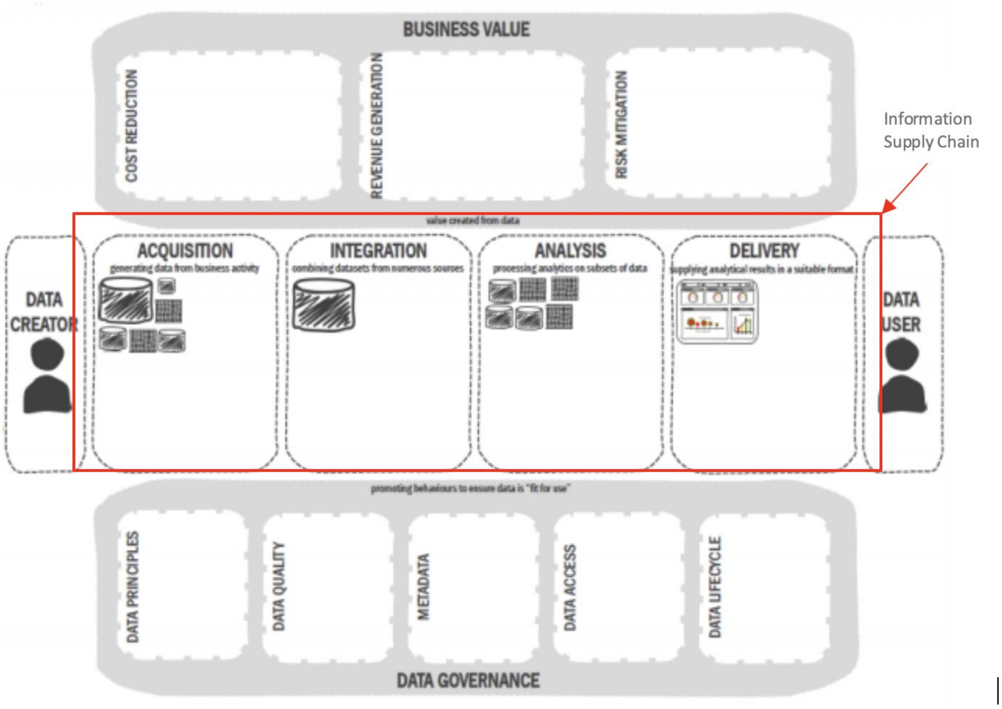


Figure 2. The Data Value Map

This study concentrates on four key components of the DVM - the ‘information supply chain’ - as a lens to examine how the various criteria impact business value specific to ML in portfolio management. In addition, the information supply chain gives a good representation of a data set life cycle within a company, making it easier to pinpoint where exactly the various criteria impact business value for portfolio management. The four key components of the DVM are now explained.

Acquisition: For many companies, the data gathered through the acquisition phase has a high degree of value (Lyko et al., 2016). The organisations that want to maximize value must take time to understand and study the value of the data they have acquired. Those institutions that do not take this time are vulnerable to costly implementation programs that will unlock very little value (PWC, 2019).

Integration: The integration component of the DVM describes the fusing of data from the different sources they were derived from. Chappell (2009) found that the combination of applications and data are focal to business value. Getting the most from this relationship is quite difficult and often requires “connecting things in intelligent ways. In other words, it requires integration” (Chappell 2009, pg.3). Data integration adds business value, as it guarantees that a data set has the same purpose and usability throughout time with all users.

Analysis: The analysis component of the DVM considers the use of analytics on data from the integration component, for informed decision making and more efficient risk discovery (Davenport and Harris 2007, Davenport et al. 2010). The importance of analysis is seen in a survey, published in MIT Sloan Management Review, where Kiron and Shockley (2011) found that companies that are better at analytics are gaining a competitive advantage. In financial organisations, data and information are at the heart of investment choices. Converting data into ‘usable information’ allows businesses and institutions to practice better decision making (Barham, 2017).

Delivery: The results generated from the previous component are examined in more detail in the final component of the DVM. The purpose of this component is to understand the needs of data users and determine how these users will receive their data (Nagle and Sammon, 2017). Understanding data users' needs can have a profound positive impact on business value. Some of the benefits to users are "better information service allowing new organization structures, more effectively coordinated/directed local autonomy and may make for more convenient office work" (Langefors 1978, p.7)

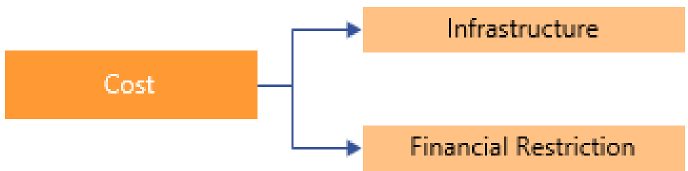
This study employs the data value map "information supply chain" as a lens to identify how the criteria the Portfolio Management industry finds necessary to implement ML impacts on business value (figure 3). Framing criteria using the data value map helps to emphasise how and where various criteria are encountered (Hewitt-Dundas, 2006), which is instrumental towards the success of technological advancements such as the implementation of machine learning in portfolio management.

The sections below are based on how the concepts identified in the literature "fit" with, and impact on, the core elements of the data value map.

BUSINESS VALUE

To successfully implement ML to automate and optimize portfolio management tasks, it is important to ensure that all necessary factors needed for ML's optimum performance, are met.

Cost



How does having the infrastructure to implement ML impact business value

Machine Learning requires the analysis of large amounts of data and needs an infrastructure to support it. Having the right infrastructure to implement ML impacts on business value for the Portfolio Management industry at the acquisition, integration, analysis, and delivery stages. Lack of data infrastructure may be a problem for some of the Portfolio Management industry as many financial institutions still use their old IT infrastructure with data silos and many legacy systems (Bhagat, 2020). Copeman and Ivanova, (2017) noted a decline in output efficiency of technology stemming from constraints on infrastructure. While there is evidence of rapid growth in machine learning throughout the finance industry, this may impose scalability challenges for teams deploying the required infrastructure to enable such advancements (Hazelwood, 2018). Furthermore, a salient feature of portfolio management companies is the impact of the massive amounts of data that is potentially available to train machine learning models. Establishments with similar scales of data have been identified as having implications that span their entire infrastructure stack (Hazelwood, 2018). Exploiting pre-existing hardware resources is a possible procedure for enabling machine learning efficiency in some organisations; however, this does involve the process of adapting or redesigning algorithms to enable effective execution (Bekkerman et al, 2011).

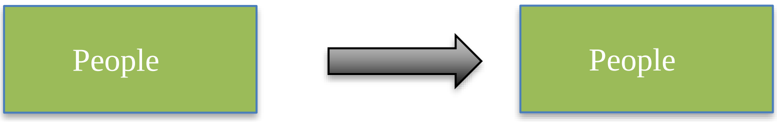
Foresight towards establishing the ideal infrastructure best caters for company goals, enables the consolation and cost minimisation of business processes, whilst simultaneously ensuring the rapid launch of future electronically based business initiatives. Robust IT infrastructure capabilities are essential for current and future electronically based processes (Weill et al., 2002), especially machine learning initiatives, to ensure optimum efficiency throughout the entire process lifecycle.

How does access to Finance impact business value

Financial restrictions impact business value for the Portfolio Management industry in all areas of the business. The existence of financing constraints is endogenous to decisions relating to innovation (Savignac, 2008; Howell, 2017; Fan et al., 2017). In a survey conducted by Sculley et al. (2018), financial limitations were identified as the most prevalent obstacle impeding the implementation of innovative projects and preventing staff from taking the lead for such initiatives. ML is an innovative procedure (Deloitte, 2018), so finance plays a large part in its success.

The portfolio management industry has been identified as one of the biggest industries making the most profit (Cooper et al, 2001), so this means there should be fewer financial strains imposed towards implementing ML and thus enhancing business value generation for the entire organisation.

People



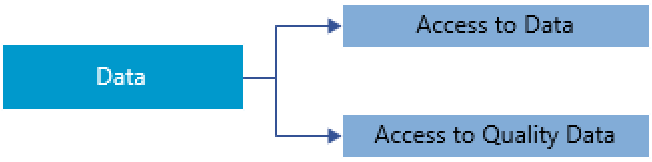
How does having People to implement ML impact business value

Discretionary fund management has more of an emphasis on the human skills and decisions than systematic fund management with its emphasis on computer implemented rules with limited human intervention. It would appear that systematic fund management would align better with ML, but the distinction between the two types of fund management is likely to become less distinct (Harvey et al., 2017). It is likely that people will continue to play a large part in portfolio management,

Having the personnel to implement ML impacts business value for the Portfolio Management industry. Attaining workers with suitability towards specified tasks have been highlighted as the 'critical index of competition' for businesses, to the extent that the development of such capacities has been given high levels of priority (Tim and Brinkerhoff, 2008). The field of ML is no different, with there being a huge demand for ML specialists (Alekseeva, 2021). Linkedin (2017) emerging US job report showed that machine learning engineers are the most sought-after profession in the US.

To provide training, in the hope of sculpting a cohort of workers better qualified for conducting tasks efficiently, requires sufficient time (Aziz and Dowling, 2019) and substantial funding (Bauer and Bender, 2004; De Bruecker et al., 2015). Furthermore, skill substitution and cross-training, which constitute solutions for some organisations to help prepare staff for certain roles, involves the risk of negative consequences on the organization's overall performance (De Bruecker et al., 2015).

Data



How does access to data impact Business Value

Access to data impacts business value for Portfolio Management at the acquisition phase. Lyko et al. (2016) outlined that the majority of data acquisitions have high velocity, high volume, and high variety but low-quality data, putting credence on the importance of having "adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value fragments of the data are actually processed by the data-warehouse analysis" (Lyko et al 2016, pg.39).

PWC (2019) noted that data can add value when it is coupled with a strategic plan. By ensuring a robust strategic led plan, organizations can link their strategy to their data, clarify what data they have access to and identify how valuable it is. Accenture (2019) detailed that companies need to become "Data Driven Enterprises", maximising the value of the data that they have access to and treating it as an asset characterized by its lineage, completeness, and quality. Data access is key to ML thriving. In some instances, it can be the key to the success of the business, "specifically, if there are increasing returns to scale or scope in data acquisition", and it is likely that early adopters or aggressive entrants can become the most competitive over the long term" (Cockburn et al., 2019, p. 118). As more and more institutions look to adopt the tools that are focal to ML advancement, the monetary incentive to obtain new data increases (FSB, 2017). From the perspective of Portfolio Management when deploying ML one of the biggest constraints to the deployment is not having access to sufficient data, but Ali et al. (2016) finds that the Portfolio Management industry is largely data-driven. As a result of this, access to data should not constitute a barrier for ML adoption in the Portfolio Management industry.

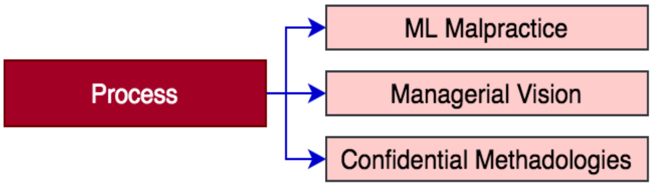
How does availability of quality data impact business value

Cooper et al. (2002) reveal that reliable financial data for portfolio managers are often unavailable during the phases of a project which involve essential decision making and are critical towards the objectives of the project. ML could simplify analysis for portfolio managers allowing them to implement more advanced investment strategies and bolster risk analysis; high-quality data is essential for this (PWC, 2018)

Availability of quality data in ML impacts business value at the acquisition and integration stage. Machine learning operations rely heavily on having access to enough high-quality data (Loshin, 2011), so the absence of this could slow down, reduce the quality, or even generate erroneous results, which will deteriorate the successful flow of all following operations and ultimately, leading to negative impacts on the overall business value.

Lyko et al. (2016) stated that for some organizations, most data is potentially of high value. "For such organizations, data analysis, classification, and packaging on very high data volumes play the most central role after the data acquisition" (Lyko et al 2016, pg.39). Companies that do not have processes for identifying and handling data quality issues integrated into their business expose risks to the usability and trustworthiness of the data, leading to negative financial impacts. This suggests that there is value in instituting processes for examining, measuring, reporting, reacting to, and controlling multiple aspects of poor data quality (Loshin, 2011).

Process



How does a strong Managerial Vision impact business value

Managerial vision impacts business value at the integration, analysis, and delivery stages. Senior management can have a large impact in unlocking the value of a company's most important programs. Chan et al. (1997) suggested that in the case of large information projects (such as ML), a company is required to create a "fit" between business strategic orientation and project strategic orientation. Managers are also expected to provide a long-term view for business value within a company. The burden of creating a strong data-driven culture within an organization falls on senior management. In most instances, organizations are lacking strong leadership in relation to data and data analysis (Pugna et al., 2019). Machine learning is an example of data analytics and due to the difficulty of implementing data analytics projects in the Financial Services sector, it is imperative that management is at the heart of its adoption.

How does Confidential Business Information impact business value

Confidential business information impacts business value for Portfolio Management at the acquisition, integration, analysis, and delivery stages. Confidential business information refers to the protection of valuable information from unauthorized parties (Almeling, 2012; Dobrusin, 2012; De Martinis et al., 2013). This practice is an issue in the Portfolio Management industry, as Sonnenburg (2007) found that the lack of open sourcing of machine learning algorithms is a major obstacle in the financial services industry and acknowledges that there is a lack of incentives for sharing ML algorithms. Although a lack of open-sourcing may be beneficial to a minority of institutions that have access to ML algorithms, it has a negative effect on the industry as a whole. Some of the main traditional benefits of open sourcing for businesses are "guaranteed support with service levels, legal indemnification and liability, warranties, regular updates and patches and no big upfront investment and significantly lower costs" (Verberg 2018, pg.2). The idea that the only benefit of open sourcing is the cost savings from subscription fees is a misleading one. Open sourcing gives institutions the licence to employ any software they need, to take on experts who can study and improve the software and thrive in a sector where there is no charge to work with the source (Verberg, 2018). These factors could have a positive impact on business value from the acquisition to the delivery stage. For Portfolio Management, the finance industry is making leaps towards a new age of higher engagement, where "financial services and fintech firms are increasingly realizing that active participation in open-source development is necessary to fully realize their goals to spur innovation, reduce development costs, and attract top talent" (Aitken et al. 2018, pg.6).

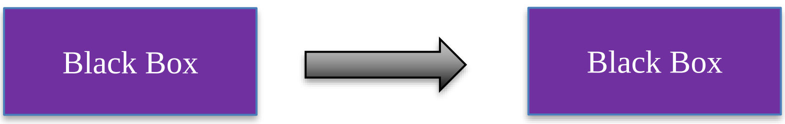
How does Machine Learning Malpractice impact business value

Specialists and researchers are trusted to push the boundaries of this cutting-edge area, but some research and practical findings can be manipulated (Ioannidis ,2005). Portfolio Managers study statistics to decide which fund to invest in; however, many statistics including the popular Sharpe ratio, which assists investors in understanding the return of an investment in comparison to its risk, are often manipulated to show more positive results (Bailey et al., 2014; Cvitanic et al, 2007; Sharpe,1994).

This directly affects the analysis aspect of the data supply chain. Social and political pressures also result in programmers using "overfitting" to exaggerate the benefits of their system. Contorting the data set in this way negatively affects the integration of information within the process. They then use the resulting Shape ratio to show greater returns on investment in a way that is mathematically negligent in order to attract investment (Harvey and Liu, 2015).

As such data users are expected to trust that the sample testing used in the creation of this process was thorough and therefore that the process will result in real gains despite their lack of understanding in the area. This can result in misfit processes being made active that defy easy understanding and can cause long term damage to the business (Bailey et al., 2014).

Black Box



What impact can Black boxes have on business value

The benefits of ML in the Analysis of data often comes at the cost of increased model complexity and a lack of explanatory insight (Van Liebergen, 2017); this is specifically presented in the dangers of black box processing. ML is vulnerable to attackers that have access to a program's learning parameters and environment. In the case of a black box, malicious inputs can be manufactured to give erroneous model outputs that appear unmodified to outside users (Papernot et al., 2017). In black box scenarios, adversarial examples may exploit the vulnerabilities in ML reinforcement learning and completely avoid detection (Sethi and Kantardzic, 2018). Adversarial examples refer to "deep neural networks have been recently found vulnerable to well-designed input samples" (Yuan et al. 2019, p.2805). Compromised processes can be exploited by bad actors which will decrease the potential business value return of the process (Papernot et al., 2016). These attacks form a distinct vulnerability within the Analysis segment of any give ML based data driven process.

FINDINGS

Through the use of the DVM, this study has found that issues of Cost and Personnel have broad reaching implications throughout the process of extracting business value from data using ML. Processes that are unavailable due to being suppressed by patents are similarly shown to have an effect throughout the information supply chain. As such they must be taken into account for every stage at which a Portfolio Manager uses or designs a ML system. Concerns around access to and quality of data are found early in the process. With quality and quantity of data forming a constraint for the overall effectiveness of a ML process. ML technology concerns such as its black box component are of major concern during the actual data analysis. While lastly, Managerial Vision and ML Malpractice are areas that take precedence during important internal processes surrounding information integration, analysis, and delivery. Visualising these criteria on the DVM helps to identify where and how these criteria impact business value for the Portfolio Management industry.

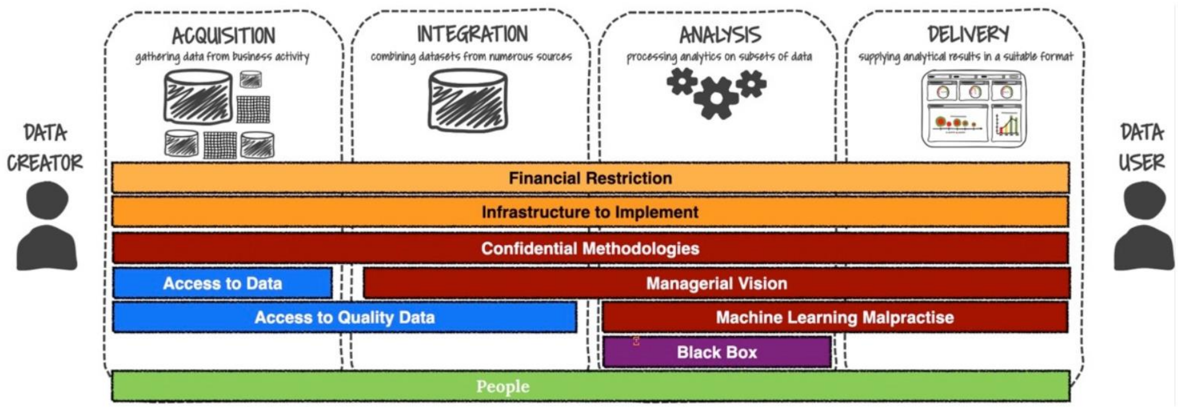


Figure 3. Impact of ML in Portfolio Management throughout the Information Supply Chain

The findings of the study show that implementing ML in Portfolio Management is becoming more and more relevant to organizations in gaining a cutting edge over their competitors. As discussed, the infrastructure is critical to the ability to extrapolate useful information. Of course, any organization looking to implement ML needs access to finance, however, the Portfolio Management industry has access to large amounts to finance so this should not be an inhibitor towards increased implementation of ML. This finance will enable the Portfolio Management industry to engage relevant personnel who can source vast quantities of highly relevant quality data. The personnel employed must exhibit leadership qualities in order to ensure that algorithms run correctly, and the organization is gaining maximum value from the process. The increase in open sourcing means that Portfolio Managers are now able to learn from others and identify what algorithms correctly interpret data relevant for their organization. Caution is required as a lack of understanding in the area can lead to malpractice; this would not be as prominent an issue if there was an increased supply of trained personnel. Better infrastructure and more trained personnel can help the PM industry prevent black box attacks.

CONCLUSIONS

This paper provides an analysis of existing literature for both academics and practitioners. It also shows the benefits of the data value map as a lens for the analysis of machine learning in portfolio management projects.

Overall, the study calls for high degrees of rigour in the specification of the implementation of Machine Learning in portfolio management. It is of utmost importance for both theory and practice to clearly understand the concepts, dimensions, and targets of this process if we are to develop a meaningful roadmap in this area. To successfully translate the hype or promises around Machine Learning technology into viable financial applications, both practical and research efforts are currently lacking.

The Data Value Map can form the basis of communication between stakeholders to ensure a holistic understanding by all. Employing the Data Value Map allows for easy comparison between the various phases, and this facilitates the formulation of questions and discussions around corrective measures that portfolio managers should consider when implementing Machine Learning in ways that are unique to their business. If a Portfolio Management firm ensures that these barriers are dealt with effectively at the appropriate stage, then the feasibility of using Machine Learning will be more widely accepted and will yield favourable results.

REFERENCES

Accenture (2019). The power of a Data Driven Enterprise. Accenture. Available at: https://www.accenture.com/_acnmedia/pdf-109/accenture-ao-dde-pov-v5.pdf

Ait Gougam, L., Chikhi, A., Mekideche-Chafa, F. (2008). Function approximation of tasks by neural networks. 6th Conference on Nuclear and Particle Physics: 443-450

Aitken, A., Hawthorn. L., Williamson, A. (2018). *Business Value of Open Source for Financial Services Firms*. Fintech Open Source Foundation: Whitepaper. Available at: https://www.finos.org/hubfs/whitepapers/FINOS-business-value-of-open-source.pdf

Alekseeva, L., Azar, J., Giné, M., Samila, S., Taska, B (2021) *The demand for AI skills in the labor market*. Labour Economics, 71: 1-41.

Ali, O., Suwabe, T., Walsh, D. (2016). *The Role of Big Data in Investing*, Goldman Sachs. Available at: https://www.gsam.com/content/gsam/global/en/market-insights/gsam-insights/gsam-perspectives/2016/big-data/gsam-roundtable.html

Almeling, D. (2012) *Seven Reasons Why Trade Secrets Are Increasingly Important*. Berkeley Technology Law Journal, 27(2): 1091-1118

Aziz, S. and Dowling, M., (2019). *Machine Learning and AI for Risk Management*. In Disrupting Finance. Palgrave Pivot, Cham: 33-50

Bailey, D., Borwein, J., Lopez de Prado, M. and Zhu, Q.J. (2014). *Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance.* Notices of the American Mathematical Society, 61(5): 458-471.

Barham, H. (2017). *Achieving Competitive Advantage Through Big Data: A Literature Review*, 2017 Portland International Conference on Management of Engineering and Technology (PICMET): 1-7

Bauer, T. and Bender, S. (2004) *Technological change, organizational change, and job turnover*. Labour Economics, 11(3): 265-291.

Bean, R, (2018) *Big Data Executive Survey 2018 Executive Summary of Findings Data and Innovation How Big Data and AI are Driving Business Innovation*, Sloan Management Review. Available at: https://sloanreview.mit.edu/article/how-big-data-and-ai-are-driving-business-innovation-in-2018/

Beel, J., Brunel, V. (2019) *Data Pruning in Recommender Systems Research: Best-Practice or Malpractice*. 13th ACM Conference on Recommender Systems. Vol. 2431

Bekkerman, R., Bilenko, M. and Langford, J. (2011) *Scaling up machine learning: Parallel and distributed approaches*. Proceedings of the 17th ACM SIGKDD International Conference, Article Number 4.

Bhagat, J. (2020) *How can banks survive and thrive in a world of automated finance*. Journal of Digital Banking. 4(3): 194-205.

Cabitza, F., Rasoini, R., Gensini, G. (2017). *Unintended consequences of machine learning in medicine*. Journal of the American Medical Association, 318(6): 517-518.

CFA (2019). *AI Pioneers in Investment Management*. Research Report CFA Institute. Available at: https://www.cfainstitute.org/-/media/documents/survey/AI-Pioneers-in-Investment-Management.ashx

Chan, Y., Huff, S., Barclay, D., Copeland, D. (1997). *Business strategic orientation, information systems strategic orientation, and strategic alignment*. Information systems Research, *8*(2): 125-150.

Chappell, D. (2009) *Creating business value through integration*. Available at: http://www.chappellassoc.com/writing/white_papers/Business_Value_of_Integration_v1.0--Chappell.pdf

Cockburn, I., Henderson, R., Stern, S. (2019) *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis*. NBER Conference on Economics of Artificial Intelligence: 115-146.

Cohen, P., Jensen, D. (1997) *Overfitting explained*. Proceedings of the Sixth International Workshop on Artificial Intelligence: 115-122

Cooper, R., Edgett, S. and Kleinschmidt, E. (2001). *Portfolio management for new product development: results of an industry practices study*. R&D Management, 31(4): 361-380.

Cooper, R., Edgett, S., Kleinschmidt, E. (2002). *Portfolio management: fundamental to new product success*. The PDMA Toolbook for New Product Development: 331-364.

Copeman, T. and Ivanova, K. (2017) *Technology Commercialization with Infrastructure Constraints*. Available at SSRN: https://ssrn.com/abstract=3063522

Corrales, D., Ledezma, A., Corrales, J. (2018). *From theory to practice: A data quality framework for classification tasks*. Symmetry, 10(7): 1-29.

Creed, A., Cotter, A., Merriman, L., McAvoy, J., O'Reilly, P. (2021). *The Delegation of Investor Decision Making: What Drives Participants in Social Trading Networks to Engage in Copy Trading*. Drake Management Review, 10(2): 4-24

Cvitanic, J., Lazrak, A., Wang. T. (2008). *Implications of Sharpe Ratio as a Performance Measure in Multi-Period Settings*. Journal of Economic Dynamics and Control. 32(5): 1622-1649.

Davenport, T., Harris, J. (2007) *Competing on analytics: The new science of winning*, Boston: Harvard Business Press.

Davenport, T., Harris, J., Morison, R. (2010) *Analytics at work: Smarter decisions, better results*, Boston: Harvard Business Press.

Deloitte (2017). Business impacts of machine learning Available at: https://www2.deloitte.com/content/dam/Deloitte/at/Documents/technology-media-telecommunications/at-tmt-predictions-2020.pdf.

Deloitte (2018). *Opening the black box Managing algorithm risks*. Deliotte Risk Advisory. Available at: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-risk-managing-algorithmic-risks-noexp.pdf

De Bruecker, P., Van den Bergh, J., Beliën, J., Demeulemeester, E. (2015). *Workforce planning incorporating skills: State of the art*. European Journal of Operational Research, 243(1): 1-16.

De Martinis, L., Gaudino, F., Respess, T. (2013) *Study on Trade Secrets and Confidential Business Information in the Internal Market. Final Study*. Available at: http://ec.europa.eu/internal_market/iprenforcement/docs/trade-secrets/130711_final-study_en.pdf

Discenza, R., Forman, J. (2007). *Seven causes of project failure: how to recognize them and how to initiate project recovery*. Paper presented at PMI Global Congress North America, Atlanta.

Dobrusin E., Krasnow R. (2012) *Intellectual Property Culture: Strategies to Foster Successful Patent and Trade Secret Practices in Everyday Business*. 2nd edition. New York: Oxford University Press

DuBois, J. (2004). *Portfolio Management Using Questionable Quality Data*. Presented at SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers.

Fan, Y., Kühn, K., Lafontaine, F. (2017). *Financial constraints and moral hazard: The case of franchising*. Journal of Political Economy, 125(6): 2082-2125.

FSB (2017). *Artificial intelligence and machine learning in financial services*. Available at: http://www.fsb.org/wp-content/uploads/ P011117.pdf

Gillespie, T. (2014). *The Relevance of Algorithms*. In Media Technologies: Essays on Communication, Materiality, and Society: 167–194.

Goldman, C. (1999) *Align drive – expert advice*. CIO Magazine, January, Vol. 5.

Harvey, C., Y. Liu. (2015) *Backtesting*, SSRN working paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2345489.

Harvey, C., Rattray, S., Sinclair, A., Van Hemert, O. (2017). *Man vs. Machine: Comparing Discretionary and Systematic Hedge Fund Performance*. Journal of Portfolio Management, 43(4): 55–69.

Hawkins, D. (2004). *The problem of overfitting*. Journal of chemical information and computer sciences, 44(1): 1-12

Hazelwood, K., Bird, S., Brooks, D., Chintala, S., Diril, U., Dzhulgakov, D., Fawzy, M., Jia, B., Jia, Y. & Kalro, A. (2018) *Applied machine learning at facebook: A datacenter infrastructure perspective.* IEEE International Symposium on High Performance Computer Architecture (HPCA), 2018. IEEE: 620-629.

Hewitt-Dundas, N. (2006). *Resource and capability constraints to innovation in small and large plants*. Small Business Economics, 26(3): 257-277.

Howell, S. (2017) *Financing innovation: Evidence from R&D grants*. American Economic Review, 107(4): 1136-64.

Ioannidis, J. (2005) *Why most published research findings are false*. PLoS Medicine, 2(8): 124.

Jonas, D.. Kock, A., and Gemunden, H. (2013). *Predicting project portfolio success by measuring management quality - A longitudinal study*. IEEE Transactions on Engineering Management 60 (2): 215–26.

Joshi, A., Wheelock, C (2018). *Storage Architecture and Implementation Considerations for AI Deployments*. PureStorage Whitepaper. Available at: https://www.purestorage.com/content/dam/pdf/en/white-papers/protected/wp-tractica-storage-for-ai-deployments.pdf

Kiron, D., Shockley, R. (2011) *Creating business value with analytics*. MIT Sloan Managemtn Review. 53(1): 57-63.

Kock, A., Gemünden, H. (2016). *Antecedents to decision-making quality and agility in innovation portfolio management*. Journal of Product Innovation Management, 33(6): 670-686.

Konstantinov, G., Chorus, A., Rebmann, J. (2020) *A Network and Machine Learning Approach to Factor, Asset, and Blended Allocation*. Journal of Portfolio Management, 46(6): 54–71.

Krause, J., Perer, A., Ng, K. (2016). *Interacting with predictions: Visual inspection of black-box machine learning models*. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems: 5686-5697.

Langefors, B. (1978) *Analysis of user needs*. In Conference of the European Cooperation in Informatics. Springer, Berlin, Heidelberg: 1-38

LaValle, S., Lesser, E., Shockley, R., Hopkins, M., Kruschwitz, N. (2011) *Big data, analytics and the path from insights to value*. MIT Sloan management review, 52(2): 21-32.

Linkedin (2017) Linkedin's 2017 us emerging jobs report. Available at: https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report

López de Prado, M., 2018. *The 10 Reasons Most Machine Learning Funds Fail*. Journal of Portfolio Management, 44(6): 120–133.

Loshin, D., 2011.*Understanding the financial value of data quality improvement*. Available at: https://www.informatica.com/downloads/1530_KnowledgeIntegFinValueDQ.pdf

Lyko, K., Nitzschke, M., Ngomo, A. (2016). *Big data acquisition*. In New Horizons for a Data-Driven Economy. Springer, Cham: 39-61

Martens, B. (2018). *The impact of data access regimes on artificial intelligence and machine learning*, European Commission JRC Digital Economy Working Paper, No. 2018-09.

Martinsuo, M., Killen, C. (2014). *Value management in project portfolios: Identifying and assessing strategic value*. Project Management Journal, 45(5): 56-70.

Michalkova, L (2018) *The Analysis of Causes of Business Financial Distress*. In Third International Conference on Economic and Business Management (FEBM 2018): 49-56

Mohajan, H. (2017). *Two criteria for good measurements in research: Validity and reliability*. Annals of Spiru Haret University. Economic Series, 17(4): 59-82.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. (2009). *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement*. Annals of internal medicine, 151: 264-269.

Mulvey, J., Liu, H. (2016) *Identifying Economic Regimes: Reducing Downside Risks for University Endowments and Foundations*. Journal of Portfolio Management, 43(1): 100–108.

Nagle, T., Sammon, D. (2017) *The Data Value Map: A framework for developing shared understanding on data initiatives*, Proceedings of the 25th European Conference on Information Systems, Guimarães, Portugal: 1439-1452.

Papernot, N., McDaniel, P. and Goodfellow, I. (2016) *Transferability in machine learning: from phenomena to black-box attacks using adversarial samples*. arXiv preprint arXiv:1605.07277.

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z., Swami, A. (2017) *Practical black-box attacks against machine learning*. In Proceedings of the 2017 ACM on Asia conference on computer and communications security: 506-519.

Perrin, S., Roncalli, T. (2019) *Machine Learning Optimization Algorithms & Portfolio Allocation*. In Machine Learning for Asset Management: New Developments and Financial Applications: 261-328.

Pugna, I., Dutescu, A., Stanila, O. (2019). *Corporate attitudes towards big data and its impact on performance management: A qualitative study*. Sustainability. 11(3): 684-709.

PWC (2018) *Asset & Wealth Management Revolution*. Available at: https://www.pwc.com/gx/en/asset-management/asset-management-insights/assets/pwc-awm-revolution-pressure-on-profitability.pdf

PWC (2019) *Creating Value for Data*. PWC. Available at https://www.strategyand.pwc.com/gx/en/insights/2019/creating-value-from-data/creating-value-from-data.pdf

Robertson, G. (2015) *To Expedite Innovation, Give Away Your Code*. Available at: https://techcrunch.com/2015/08/04/to-speed-up-innovation-give-away-your-code/

Ryll, L., Barton, M., Zhang, B., McWaters, R., Schizas, E., Hao, R., Bear, K., Preziuso, M., Seger, E., Wardrop, R., Rau, P. (2020). *Transforming Paradigms: A Global AI in Financial Services Survey*. Available at: https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/2020-ccaf-ai-in-financial-services-survey.pdf

Savignac, F. (2008) *Impact of financial constraints on innovation: What can be learned from a direct measure*. Economics of Innovation and New Technology, 17(6): 553-569.

Sculley, D., Snoek, J., Wiltschko, A., Rahimi, A. (2018) *Winner's curse? On pace, progress, and empirical rigor*. International Conference on Learning Representations Workshop track

Sethi, T. and Kantardzic, M. (2018) *Data driven exploratory attacks on black box classifiers in adversarial domains*. Neurocomputing, 289: 129-143.

Sharpe, W. (1994) *The Sharpe ratio*. Journal of Portfolio Management, 21(1): 49-58.

Sonnenburg, S., Braun, M, Ong, C., Bengio, S., Bottou, L., Holmes, G., LeCun, Y., Muller, K., Pereira, F., Rasmussen, C., Ratsch, G. (2007) *The need for open source software in machine learning*. Journal of Machine Learning Research, 8: 2443-2466.

Stentoft, L., Wang, S. (2020) *Consistent and Efficient Dynamic Portfolio Replication with Many Factors*. Journal of Portfolio Management, 46(2): 79–91.

Stork, D. (2000). *Open data collection for training intelligent software in the open mind initiative*. In Proceedings of the Engineering Intelligent Systems (EIS2000).

Tim M. and Brinkerhoff, R.O. (2008). *Courageous Training: Bold Actions for Business Results*. Colorado, Berrett-Koehler Publishers.

Van den Steen, E. (2005). *Organizational beliefs and managerial vision*. Journal of Law, Economics, and Organization 21(1): 256-283

Van Liebergen, B. (2017). *Machine learning: A revolution in risk management and compliance*. Journal of Financial Transformation, 45: 60-67.

Vanson Bourne (2015). *The State of Big Data Infrastructure: Benchmarking global Big Data users to drive future performance*. Available at https://docplayer.net/2116103-The-state-of-big-data-infrastructure-benchmarking-global-big-data-users-to-drive-future-performance.html

Verberg, J. (2018). *The Benefits of Commercial Open Source*, Bloomreach. Available at: http://go.bloomreach.com/rs/243-XLW-551/images/Hippo-WP-commercial-open-source-benefits.pdf

Von Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A. (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature search process.* ECIS 2009 Proceedings. 161.

Voss, M., Kock, A. (2013). *Impact of relationship value on project portfolio success - Investigating the moderating effects of portfolio characteristics and external turbulence*. International Journal of Project Management 31 (6): 847–61.

Walter, M., Zimmermann, J. (2016). *Minimizing average project team size given multi-skilled workers with heterogeneous skill levels*. Computers & Operations Research, 70: 163-179.

Webster, J. & Watson, R. (2002) *Analyzing the past to prepare for the future: Writing a literature review*. MIS quarterly, 26(2): xiii-xxiii.

Weill, P., Subramani, M., Broadbent, M. (2002) *IT infrastructure for strategic agility*. MIT Sloan Management Review. 44(1): 57-65

Yuan, X., He, P., Zhu, Q., Li, X. (2019). *Adversarial Examples: Attacks and Defenses for Deep Learning*. IEEE Transactions on Neural Networks and Learning Systems, 30(9): 2805-2824