PRIFYSGOL BANGOR / BANGOR UNIVERSITY

PRIFYSGOL
BANGOR
UNIVERSITY

The Role of Data Analytics within Operational Risk Management: A Systematic Review from the Financial Services and Energy Sectors

Cornwell, Nikki; Bilson, Christopeher M; Gepp, Adrian; Stern, Steven; Vanstone, Bruce J

Journal of the Operational Research Society

DOI:
https://doi.org/10.1080/01605682.2022.2041373

Published: 01/01/2023

Publisher's PDF, also known as Version of record

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Cornwell, N., Bilson, C. M., Gepp, A., Stern, S., & Vanstone, B. J. (2023). The Role of Data Analytics within Operational Risk Management: A Systematic Review from the Financial Services and Energy Sectors. *Journal of the Operational Research Society*, *74*(1), 374-402. https://doi.org/10.1080/01605682.2022.2041373

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

16. Apr. 2024

Journal of the Operational Research Society

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjor20

The role of data analytics within operational risk management: A systematic review from the financial services and energy sectors

Nikki Cornwell, Christopher Bilson, Adrian Gepp, Steven Stern & Bruce J Vanstone

To cite this article: Nikki Cornwell, Christopher Bilson, Adrian Gepp, Steven Stern & Bruce J Vanstone (2022): The role of data analytics within operational risk management: A systematic review from the financial services and energy sectors, Journal of the Operational Research Society, DOI: 10.1080/01605682.2022.2041373

To link to this article: https://doi.org/10.1080/01605682.2022.2041373

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 27 Feb 2022.

Submit your article to this journal

Article views: 1435

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjor20

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2022.2041373

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

The role of data analytics within operational risk management: A systematic review from the financial services and energy sectors

Nikki Cornwell[a], Christopher Bilson[a], Adrian Gepp[a], Steven Stern[a] and Bruce J Vanstone[a,b]

[a]Bond University, Gold Coast, Australia; [b]Bangor University, Bangor, United Kingdom of Great Britain and Northern Ireland

ABSTRACT

Operational risks are increasingly prevalent and complex to manage in organisations, culminating in substantial financial and non-financial costs. Given the inefficiencies and biases of traditional manual, static and qualitative risk management practices, research has progressed to using data analytics to objectively and dynamically manage risks. However, the variety of operational risks, techniques and objectives researched is not well mapped across industries. This paper thoroughly reviews the emerging research area applying data analytics to operational risk management (ORM) within financial services (FS) and energy and natural resources (ENR). A systematic literature search resulted in 2,538 publications, from which detailed bibliometric and content analyses are performed on 191 studies of relevance. The literature is classified using a novel multi-layered framework, informing critical analyses of the analytics techniques and data employed. Five core themes emerge, relevant to practitioners, researchers, educators and students across any sector: risk identification, causal factors, risk quantification, risk prediction and risk decision-making. Generally, ENR studies focus on identifying causal factors and predicting specific incidents, whereas FS applications are more mature surrounding risk quantification. To conclude, the comprehensive review reveals areas where further research is needed to advance ORM within and beyond FS and ENR, in pursuit of improved decision-making.

ARTICLE HISTORY
Received 8 December 2020
Accepted 5 February 2022

KEYWORDS
Analytics; risk; decision support systems; finance; energy; natural resources

1. Introduction

From the mid-1990s, disruption to the economic, financial and social environment from financial crises, corporate scandals, technological advancements and catastrophic events shifted the practice of risk management from a silo basis to an enterprise-wide approach, referred to as enterprise risk management (ERM) (Eryilmaz, 2018). ERM is "a systematic and integrated approach to the management of the total risks that a company faces" (Dickinson, 2001, p. 360). Standardised ERM frameworks, including by COSO (2017) and ISO (2018), have seen risk management elevated beyond the financial services (FS) sector (e.g. banks, insurance and superannuation) to energy and natural resources (ENR) (e.g. oil and gas, mining and power utilities), healthcare and not-for-profits (Buehler et al., 2008). The frameworks consolidate the core risk management activities of identification, analysis, evaluation, treatment, communication, monitoring and reporting. These activities traditionally involve manual processes, such as periodic and sample-based audits, assurances, likelihood-severity risk matrices and controls testing. However, the reactive and subjective nature of these approaches is limiting effective risk management (McKinsey & Company, & Operational Riskdata eXchange Association, 2017). Bromiley et al. (2015) and Aven (2016) provide more comprehensive critiques of current practice. The COVID-19 pandemic has highlighted these weaknesses, showcasing how fast-paced and pervasive the consequences of risks can be (Evans, 2020; McKibbin & Fernando, 2021).

In response, risk management is evolving to harness the value of data analytics to gain timely insights that inform risk-preventative policies, procedures, controls and early identification, ultimately reducing the frequency and severity of operational loss events. This advancement reflects the increasing ubiquity of artificial intelligence in the prevailing fourth industrial revolution (Akter et al., 2022). Data analytics involves analysing varied data to gain insights that support decision-making, and ranges from simple to advanced (for details on statistical learning see Venables and Ripley (2002) or Hastie et al. (2009), and Chollet (2018) for machine and deep learning). Risk professionals expect data analytics to transform the discipline from three key perspectives. First, an organisation's internal data,

CONTACT Adrian Gepp adgepp@bond.edu.au

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

supplemented with external data, will offer more scientific and population-based risk assessments, as compared to existing qualitative and sample-based approaches that are biased by individuals' experiences, perceptions and tolerances (Aven & Flage, 2020; Bromiley et al., 2015). Second, data-driven approaches allow for continuous scanning of the changing risk environment, compared to irregular and static assessments with manual approaches (Peters et al., 2018). Third, advanced analytics' predictive power and ability to model complex relationships provides a forward-looking approach (Aven, 2016). This will enable risk management to become a "valued component of decision-making" (Peters et al., 2018, p. 7) and drive revenue, rather than being defensive and compliance-focussed. Nateghi and Aven (2021) provide additional commentary on data analytics' contribution to risk analysis.

Research applying data analytics to risk management increased from the 1980s, and accelerated from 2012 (Aven & Flage, 2020). The application to financial risk management is prevalent, with studies estimating credit scores and default probabilities, and predicting bankruptcy (Leo et al., 2019). Non-financial risks constitute a similarly large portion of studies. Of these, global, country- or community-wide risks are commonly researched, including natural disasters, road traffic and transport risks, medical diagnosis and disease prediction (Araz et al., 2020; Choi & Lambert, 2017). Numerous studies examine operational (non-financial) risks at an organisational level, defined as the "risk of loss resulting from inadequate or failed internal processes, people and systems or from external events" (Basel Committee on Banking Supervision, 2006, p. 144). Despite slight variations by industry, operational risks generally include workplace safety; system and equipment disruptions or failures; supply chain risks; fraud and money laundering; and improper or illegal business practices. This categorisation is supported by the loss event taxonomy for banks in Basel II (Basel Committee on Banking Supervision, 2006, pp. 305–307).

Contrary to financial risk management, operational risk management (ORM) is a new area with relatively rudimentary and immature tools (Peters et al., 2018). Major operational loss events (e.g. Barings Bank rogue trader incident (Shevchenko, 2015)) and regulatory enquiries (e.g. Australian Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry (Hayne, 2019)) have raised the profile of non-financial risks. Resulting regulations have forced organisations' attention to non-financial risks (Peters et al., 2018), which the COSO and ISO risk-agnostic ERM frameworks broached. This has been exemplified in banking and aviation, touting these heavily regulated industries as having some of the most advanced risk management (PwC, 2017; Sjöblom et al., 2013). Similarly, operational risk and environmental, social and governance (ESG) regulatory changes (e.g. European Commission Non-Financial Reporting Directive (2014)) are affecting other industries, including mining (KPMG, 2021). Given ORM's infancy and the outlook of ongoing regulatory change across industries, it is timely to take stock of existing innovations, extrapolate previous works to less mature industries undergoing transformation and provide a clear trajectory for researchers and industry to further enhance ORM. This review focusses on the application of data analytics to ORM in the FS and ENR sectors, which are disparate in ORM maturity and cover a range of business processing-, system- and safety-based operational risk events across largely online versus physical environments.

This review uniquely compares findings from two sectors. While previous reviews evaluate the use of data analytics to manage a specific operational risk event or in a specific industry (Table 1), none comprehensively review operational risk holistically across a sector. Rather, they focus on a single operational risk (e.g. supply chain risk), or review data analytics applied to a single industry from either a broad optimisation (not risk management-specific) or general risk management perspective (not ORM-specific). Given these limitations, the diversity of ORM literature utilising data analytics is not well mapped regarding the breadth of risks analysed and the techniques and data used, and the similarities and differences between industries remain unknown. Yet, consistent with the ERM paradigm, it is important that operational risks are considered collectively. Additionally, industry-agnostic ERM frameworks and risk management professional bodies (e.g. International Actuarial Association, 2021; Risk Leadership Network, 2021) have highlighted the value of extrapolating risk management beyond single industries to learn from other sectors, further motivating this review's comparison between FS and ENR.

Overall, this paper aims to investigate how and what data and analytics have been used to manage operational risks across the FS and ENR sectors, so to identify avenues to advance ORM within these sectors. In doing this, the paper makes three key contributions. First, a novel multi-layered classification framework is developed. It provides a much-needed mechanism to meaningfully segment and map the literature, in a way that is consistent with fundamental risk management and data science principles to ensure readers from both disciplines benefit. The framework scaffolds critical analyses of

Table 1. Comparison of previous literature review papers.

Review paper	Risk event	Industry	Systematic	Key limitation
Operational risk event-specific				
Workplace safety				
Araz et al. (2020)	External natural or man-made disasters	Various	✗	
Shayboun et al. (2020)	Occupational accidents	Industrials – Construction	✗	Only 7 papers reviewed
Supply chain risk				
Qazi et al. (2015)	Supply chain risk	Various	✓	
Vishnu et al. (2019)	Supply chain risk	Various	✓	
da Silva et al. (2020)	Supply chain risk	Various	✓	
Fagundes et al. (2020)	Supply chain risk	Various	✓	
Hosseini and Ivanov (2020)	Supply chain risk	Various	✓	Only Bayesian Network applications reviewed
Fraud and money laundering				
Jans et al. (2010a)	Fraud	Various	✗	Papers up to 2009 reviewed
Ngai et al. (2011)	Fraud	Various	✓	1997-2008 papers reviewed
Palshikar and Apte (2013)	Money laundering	FS – Banking	✗	
Industry-specific				
FS				
Aziz and Dowling (2019)	Risk management generally (credit, market, liquidity and operational risks)	FS – Banking	✗	
Dicuonzo et al. (2019)	Risk management generally	FS – Banking	✗	
Leo et al. (2019)	Risk management generally (credit, market, liquidity and operational risks)	FS – Banking	✗	
Lacković et al. (2020)	Risk management generally	FS – Banking	✗	Discussion paper
ENR				
Shafiee et al. (2019)	Fault diagnosis, early warning detection and performance optimisation	ENR – Upstream oil and gas	✓	Business enhancement / optimisation focus, not risk management
Wu et al. (2019)	Fault diagnosis, early warning detection and performance optimisation	ENR – Nuclear	✗	Business enhancement / optimisation focus, not risk management
Industrials				
Akinosho et al. (2020)	Resource planning, risk management and logistics	Industrials – Construction	✓	Only deep learning applications reviewed; Not solely risk management focussed
Hegde and Rokseth (2020)	Risk management generally	Industrials	✓	
Risk management generally (i.e. not limited to operational risk, nor an industry)				
Choi and Lambert (2017)	Risk management generally	Various	✗	Only articles published in *Risk Analysis* reviewed
Choi et al. (2017)	Risk management generally	Various	✗	Only articles published in *IEEE Transactions* reviewed
Mišić and Perakis (2020)	Risk management generally	Various	✗	Business enhancement / optimisation focus
Nateghi and Aven (2021)	Risk management generally	Various	✗	Discussion paper – not intended to be a review of data analytics techniques for risk analysis

the large variety of analytics techniques, variables and data sources used across studies. These results improve researchers' and practitioners' understanding of the value of different techniques and data types in a risk context, promoting future research. Second, to our knowledge, this is the first systematic review and bibliometric analysis of studies applying data analytics to the array of operational risks across the entirety of the distinct FS and ENR sectors. This provides objective insights into the field's progression across the two sectors. Third, reviewing the research across FS and ENR – disparate in ORM maturity and operating environments – reveals where there are differences in their focus, as well as similarities in how data analytics can be applied to ORM. Five core research themes gleaned from the literature encapsulate these similarities generalisable across sectors, providing a platform for industries with less mature ORM practices to advance.

Section 2 outlines the systematic methodology, before the results and novel classification framework are presented in Section 3. Five core research themes are then derived in Section 4, before concluding in Section 5.

2. Systematic literature review methodology

This review focusses on FS and ENR for several reasons. Risk management has long been grounded in FS, with researchers and practitioners largely concentrating on financial risks given they are more easily quantified and directly impact financial institutions, critical to economic stability (Buehler et al., 2008). Stringent and internationally-coordinated

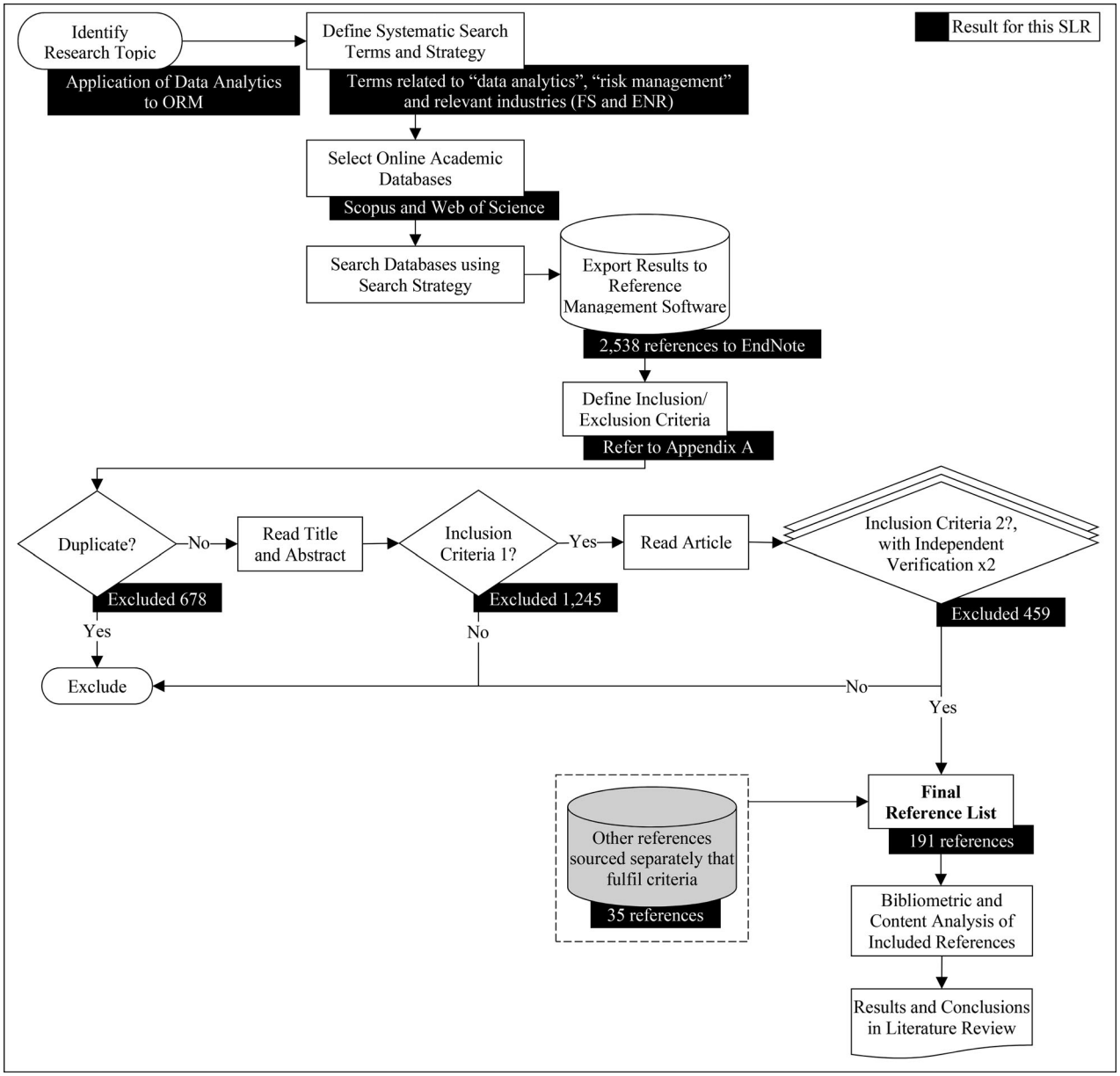

Figure 1. Process map for SLR methodology.

regulation (e.g. Basel Accords (Basel Committee on Banking Supervision, 2006, 2010)) has led FS to relatively advanced risk management (PwC, 2017). Thus, it is valuable to review data-driven ORM developments within the mature FS sector. However, FS organisations operate in mostly online and transaction-based environments.[1] Hence, FS is contrasted with a sector whose operating environment is characterised by manual labour and physical asset-intensiveness, meaning workplace safety and asset management are critical. ENR is selected to broaden the review in this respect. Compared to other safety-centric industries, mining and electricity services report high fatality rates (Safe Work Australia, 2020), which is poignant given the global oil and gas exploration and production industry is the third largest by revenue (IBISWorld, 2021).

A systematic literature review (SLR) is employed to ensure objectivity and reproducibility (Figure 1) (Linnenluecke et al., 2020). The search strategy includes terms relating to the topic's key aspects – data analytics and risk management – and the focus sectors. The terms reflect typical terminology used in each sector, derived from preliminary reading and discussions with academic and industry experts (e.g. Hegde & Rokseth, 2020; KPMG US & The Risk Management Association, 2018). For example, the risk management-related search terms are not limited to operational risks since such terminology tends to be specific to FS. The search strategy developed and validated by all authors is *("data analy*" OR "machine learning" OR "big data" OR "artificial intelligence" OR "business intelligence" OR "data min*") AND ("risk manag*" OR "risk analy*" OR "risk framework" OR "risk decision" OR "operational risk") AND ("financial services" OR bank* OR "financ* industry" OR "financ* sector" OR "energy" OR "mining")*. The largest databases containing

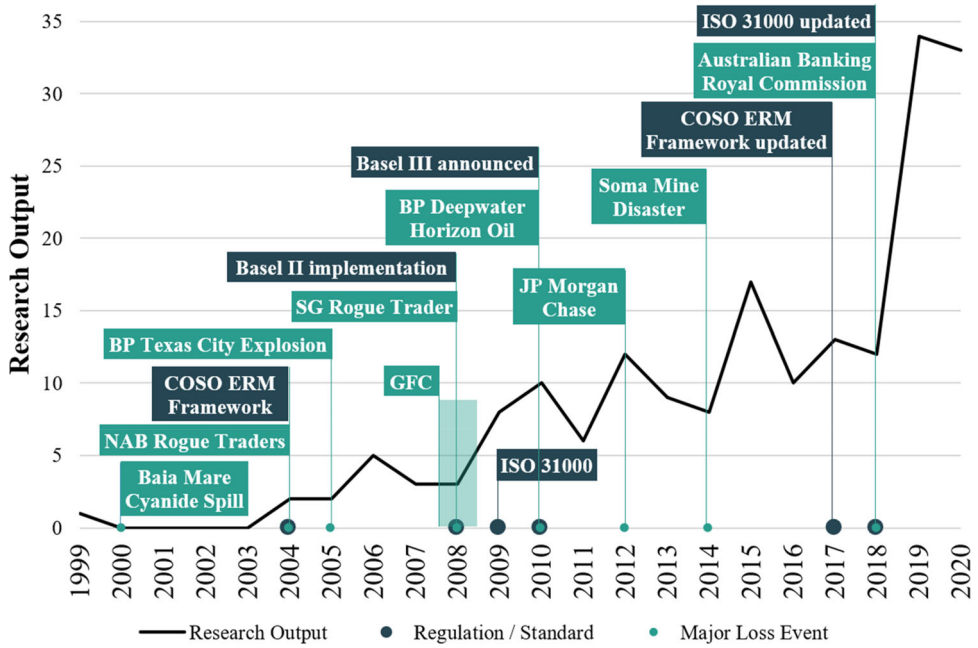

Figure 2. Annual output of research applying data analytics to ORM, overlaid with key regulatory and loss events.
Note: The regulation and standard information (COSO, 2020; ISO, 2018; Kaplan Higher Education, 2019; Standards Australia, 1995) and major loss events (Centers for Disease Control and Prevention, 2017; Kaplan Higher Education, 2019; Thomson Reuters, 2010) are obtained from various sources.

peer-reviewed literature in the field, Scopus and Web of Science, are searched and 2,538 references retrieved.[2]

A two-staged cleaning process is performed, following inclusion and exclusion criteria (Appendix A). Three authors independently verify this process with discrepancies re-evaluated according to the criteria, before reaching a consensus. This results in 156 references. An additional 35 references from Google Scholar or cited by another article are included as they fulfil the inclusion criteria. A final total of 191 references are analysed.

Bibliometric analysis is conducted using the *Bibliometrix* R package (Aria & Cuccurullo, 2017) on 186 references – five are not available through the package. Content analysis and manual classification is performed on all 191 references.

3 Results

3.1. Bibliometric analysis

Research applying data analytics to ORM across FS and ENR has progressively increased since 1999 and more than doubled from 2018 (Figure 2). This trajectory was likely motivated by major regulatory changes and loss events triggering compliance obligations and foregrounding the importance of effective ORM. For example, the surge from 2008 to 2010 followed the introduction of Basel II and new standards prioritising systematic risk management across industries (COSO, 2004; ISO, 2009). Regulators and practitioners also expanded their focus from financial to operational risks following the Global Financial Crisis (GFC) and other major loss events (Peters et al., 2018). Similarly, the rise from 2018 aligns to COSO and ISO framework updates, impending regulatory emphasis on ESG (Hayne, 2019) and the data-driven transformation megatrend (World Economic Forum, 2021). Despite recent activity, the area remains in early development with a large proportion of conference papers (40.3%) (Table 2) and only six publications with over 100 citations. The most highly cited publication is Leveson (2004), detailing Systems-Theoretic Accident Model and Processes (STAMP).

3.2. Content analysis

Throughout the manual content analysis, a new framework for classifying the literature applying data analytics to ORM across FS and ENR is developed (Figure 3). It applies to studies developing or implementing a specific model or framework; review papers are considered separately. It extends Aven's (2016) dichotomous classification of the risk field to neatly categorise and conceptualise the expansive area. The framework's five layers represent characteristics fundamental to risk and data science and are critical to understand and differentiate between the core objective and features of studies. The first layer is the *study type* – an empirical investigation or development of a generic framework, theory or model (Aven, 2016). The second layer – *risk perspective* – indicates the number of operational risks considered and the level of detail in which a study views them. It is akin to the silo versus enterprise-wide concepts introduced, yet is a three-level factor

Table 2. Most Frequently Published Reference Types and Journals.

Frequency of reference types		Frequently published journals		
Reference Type	Frequency	Journal	Frequency	Rank[a]
Journal Article	89	Risk Analysis	10	- (Q1)
Conference Paper	75	Journal of Operational Risk	7	C (Q4)
Review Paper	7			
Book Chapter	6	Computers in Industry	3	- (Q1)
Proceedings	3	IEEE Access	3	- (Q1)
Book	4	Journal of the Operational Research Society	3	A (Q2)
Editorial	2			
	186	Reliability Engineering and System Safety	3	- (Q1)
		Safety Science	3	A (Q1)
		Decision Support Systems	2	A* (Q1)
		Expert Systems with Applications	2	C (Q1)
		Process Safety and Environmental Protection	2	- (Q1)
		Risks	2	B (Q2)

bold value signifies the total.
[a]The journal ranks were sourced from the 2019 Australian Business Deans Council Journal Quality List published on 6 December 2019 (Australian Business Deans Council, 2019), as well as SCImago Journal Rankings as of July 2020 in brackets (SCImago, n.d.).

– *micro*, *multi-risk* and *macro*. *Micro* studies consider a single, highly specific risk event (e.g. floor water inrush in a mine (Wu et al., 2011)). *Macro* studies aggregate across an organisation or business line (i.e. the category of operational risk is considered, rather than individual events) (e.g. a business line's total dollar-value of operational losses over a given period (Mittnik & Starobinskaya, 2010)). *Multi-risk* considers multiple specific events across an organisation in detail (e.g. payment failures, financial exposure errors and regulatory/legal non-compliances in a bank (Sanford & Moosa, 2015)). The third layer – *analytics objective* – describes the objective of the analytics techniques concerning the risk management process – descriptive, diagnostic, predictive or prescriptive (Gudivada, 2017). Descriptive and diagnostic analytics are reactive, backward-looking analyses, providing insight into what happened and why. Predictive analytics are proactive and forward-looking and aid decision-making. Prescriptive analytics extend this, aiming to prescribe the optimal decisions. The final two layers relate to research methodology – *analytics techniques* and *data inputs* (discussed in Sections 3.2.2 and 3.2.3).

Figure 4 visualises the literature as per the framework,[3] and the following sub-sections analyse the findings by layer.

3.2.1. Overview of study types and risk perspectives

Excluding academic and industry reviews, empirical studies and generic framework developments are represented equally. This reflects the main tasks in the risk field: (i) using "risk assessments and risk management to study and treat the risk of specific activities" and (ii) performing "generic risk research and development, related to concepts, theories, frameworks … and models" (Aven, 2016, p. 1).

When further partitioned by risk perspective, approximately 40% of studies are empirical investigations into *micro* risks in specific worksites, business lines or organisations. Within ENR, equipment failures and safety-related loss events are most frequently researched (Figure 5). Information security, cyber attacks and fraud are the *micro* risk events of focus in FS. Given the specificity of the risks and contexts in *empirical-micro* studies, the methodologies and findings lack generalisability to other risks and organisations. Research developing generic frameworks to manage *micro* risks (22%) seeks to overcome this, but findings remain limited to specific risks.

Researchers have sought to understand the drivers and interconnectivity of operational risk events by incorporating multiple risks and technical, organisational, social and environmental factors. Such *multi-risk* perspectives are most prevalent in ENR safety research, with some empirical investigations (4%) into real-time monitoring and communication systems for incidents on mine sites damaging people, property or the environment (e.g. Haustein et al., 2008; Sanchez-Pi et al., 2015). Yet, reflecting its conceptual stage of development, generic frameworks dominate *multi-risk* research (e.g. Moura et al., 2017; Pence et al., 2019), constituting one-fifth of the field. This represents a shift from reactive to proactive, targeted management. Reason's (1990) system approach to human error and Swiss-cheese model, and Rasmussen's (1997) socio-technical risk analysis (SoTeRiA)[4], provide the theoretical foundation to the perspective. Viewing operational risk events in detail, yet in the whole, interconnected organisation, aims to uncover the "upstream systemic failures" which create conditions inducing human errors (Reason, 2000, p. 768).

Fewer studies (15%) take a *macro* risk perspective, evenly split across study types. The majority of these studies involve estimating banks' operational risk capital (e.g. Chavez-Demoulin et al., 2016; Dutta & Perry, 2006), motivated by regulation prescribing measurement methodologies[5] that aggregate operational risk to an organisational level (Basel

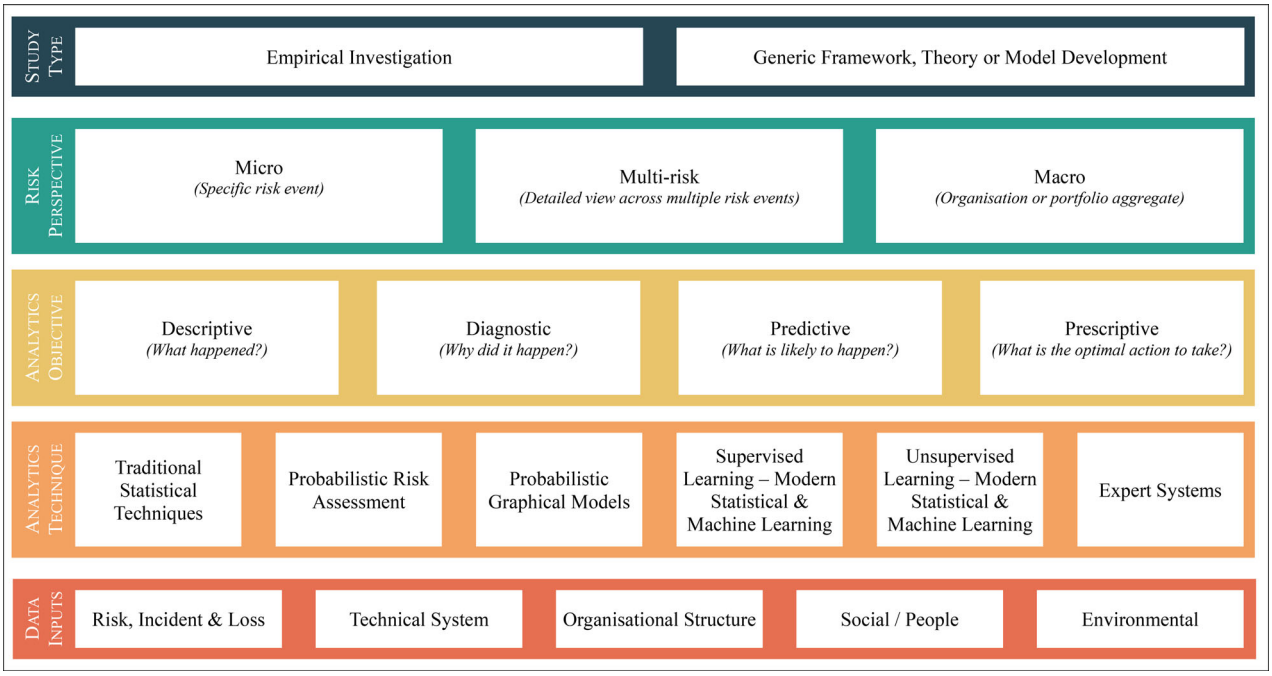

Figure 3. Classification framework for literature applying data analytics to ORM.

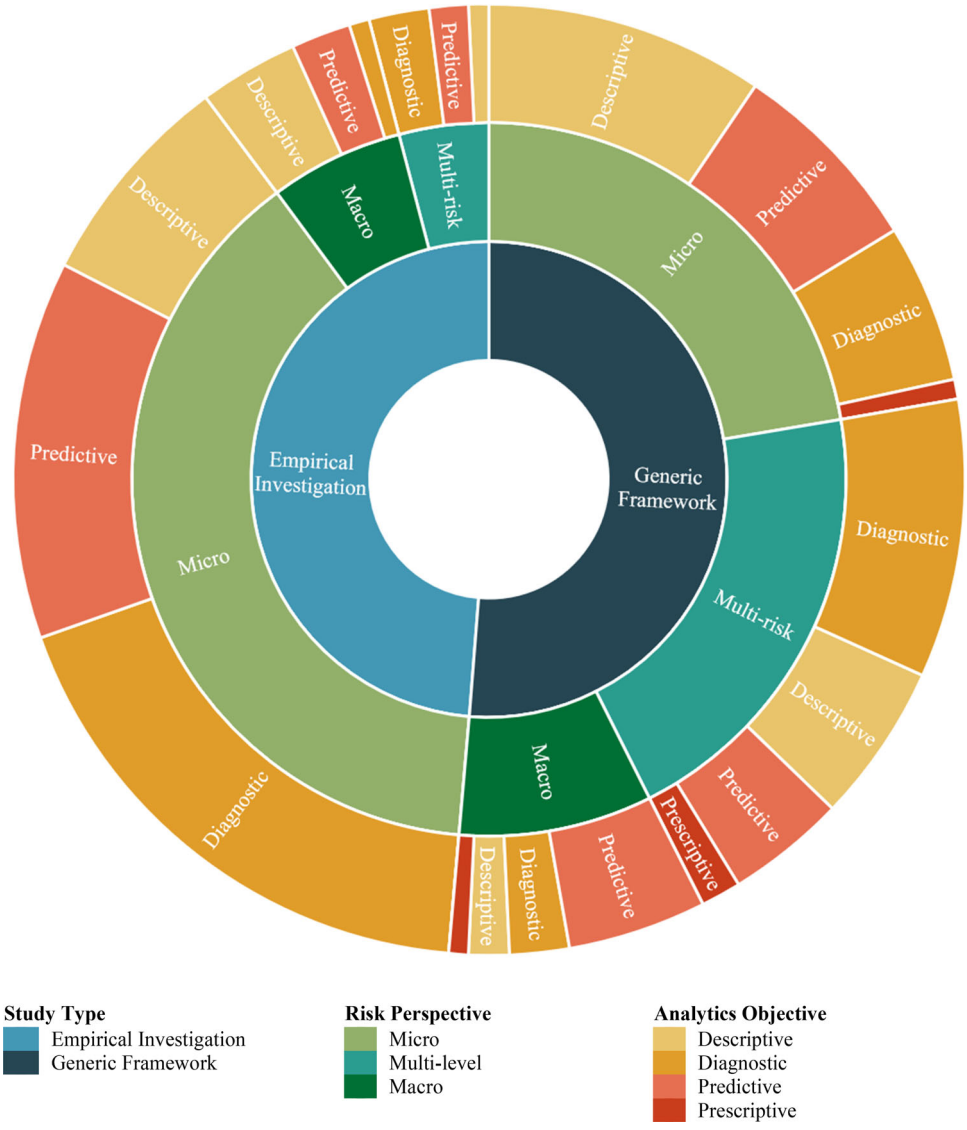

Figure 4. Breakdown of literature applying data analytics to ORM in FS and ENR as per classification framework.
Note: Figure 4 is based on the classification framework assignment of 148 references. The remaining 43 references from the 191 selected for the SLR were not included as they were not methodology papers that developed or implemented a specific model or framework, but rather they were academic or industry review papers and theoretical articles.

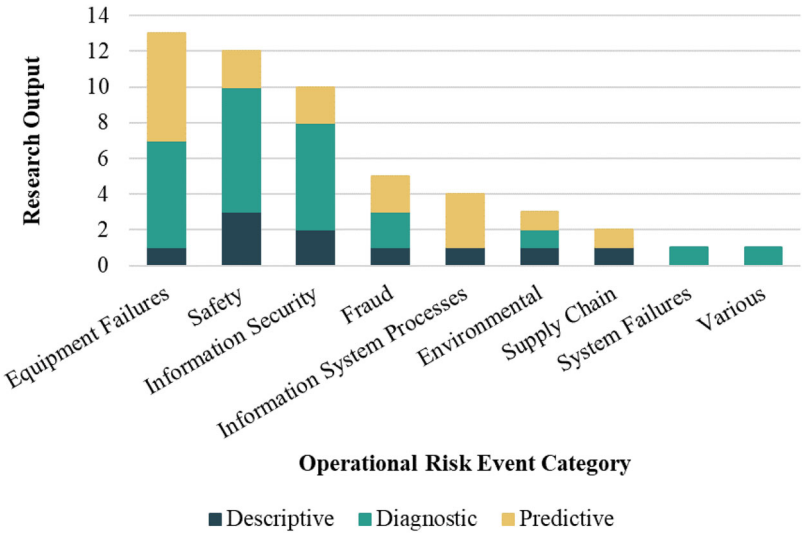

Figure 5. Frequency of empirical studies with a micro risk perspective by operational risk event category.

Committee on Banking Supervision, 2006). Other *macro-empirical* studies calculate an organisation's overall risk comparative to others in the industry (e.g. Hajakbari & Minaei-Bidgoli, 2014; L. Wei et al., 2019). Unlike *micro* and *multi-risk* perspectives, *macro* approaches do not provide visibility of the collection of operational risks facing an organisation, their underlying drivers, nor inter-relationships. This lack of interpretability restricts proactive, targeted ORM.

3.2.2. Critical analysis of analytics techniques

Techniques across various model families are used (Figure 3). Table 3 outlines their frequency, overall and by risk perspective. The categorisation is based on common statistical/data science (e.g. Hastie et al., 2009) and quantitative risk modelling (McNeil et al., 2015; Modarres et al., 2016) technique taxonomies.

Modern statistical and machine learning techniques are employed most (65%), consisting of supervised (37%) and unsupervised (28%) approaches. This broad model family refers to algorithms based on statistical learning theory that analyse and learn patterns in historic (and typically higher-dimensional) data to make inferences or predictions about unseen (future) data (Hastie et al., 2009). Supervised learning, unlike unsupervised, is guided by a dependent variable and generally lends itself to prediction tasks. Decision trees and artificial neural networks (ANNs) are used extensively to predict the occurrence, type or severity of *micro* risks, particularly in ENR. For example, Mazumder et al. (2021) compare the performance of tree-based methods at predicting oil and gas pipeline failures based on equipment and environment specifications; Marquez et al. (2020) predict degradation of LNG tank pumps to inform preventative maintenance using an ANN on sensor data. In contrast, unsupervised techniques use only independent variables to organise or cluster data for descriptive or diagnostic analyses. Association rule mining is prevalent to extract frequent co-occurrences from incident data, indicative of common incident precursors or causes that can inform prevention strategies (e.g. Abbass et al., 2020; Wu et al., 2015). Natural language processing (NLP) is often used to gain insights from unstructured risk data but are limited by linguistic ambiguity (Leidner & Schilder, 2010). Applications include identifying key risks (e.g. Chu et al., 2020) and data pre-processing to extract relevant factors from free-text reports (e.g. Pence et al., 2020).

Traditional statistical techniques are applied in approximately 15% of the literature. Rather than identifying generalisable predictive patterns as in machine learning, traditional approaches draw population inferences from a sample based on probability theory (Bzdok et al., 2018). These methods are dominant in banking, motivated by the Basel II operational risk capital requirements. Underpinned by the loss distribution approach (LDA), parametric distribution fitting or semi-parametric extreme value theory are primarily used to parameterise loss frequency and severity distributions; copula functions often define the dependencies between business lines and risk types; and Monte Carlo simulation is typically used to subsequently estimate the value-at-risk (capital level) (Dutta & Perry, 2006). These techniques rely on historic internal loss data, encompassing the loss amount, business line, risk type and year, but they generally do not consider related operational factors. This reflects the high-level, descriptive and backward-looking nature of these techniques, compared with the detailed predictive orientation of modern statistical and machine learning.

Within the family of probabilistic graphical models, Bayesian networks represent a substantial portion (14%) of the data-driven ORM research. They

Table 3. Frequency of analytics techniques used in literature applying data analytics to ORM by risk perspective.

			Risk perspective		
Model family	Technique	TOTAL	Micro	Multi-risk	Macro
Supervised learning – modern statistical & machine learning		60	50	7	3
	Decision Tree	17	12	2	3
	Artificial Neural Networks	15	13	2	
	Logistic Regression	7	6	1	
	K-Nearest Neighbours	5	5		
	Random Forest	5	5		
	Support Vector Machines	5	3	2	
	Evolutionary Algorithms	3	3		
	Fuzzy Decision Tree	2	2		
	Least Squares Regression	1	1		
Unsupervised learning – modern statistical & machine learning		45	25	15	5
	Association Rule Mining	19	11	7	1
	Natural Language Processing	9	1	5	3
	Clustering – unspecified	6	4	1	1
	Clustering – K-Means	5	3	2	
	Process Mining	3	3		
	Fuzzy Clustering	2	2		
	Self-Organising Maps	1	1		
Traditional statistical techniques		25	9	2	14
	Parametric Distribution Fitting	8	2		6
	Monte Carlo Simulation	5	1		4
	Fuzzy Theory / Rough Sets	4	4		
	Extreme Value Theory	3		1	2
	Other	3	2		1
	Copula Method	2		1	1
Probabilistic graphical models	Bayesian Networks	23	8	8	7
Probabilistic risk assessment		6	2	4	
	Event Trees	3	1	2	
	Fault Trees	3	1	2	
Expert systems[a]	Primarily rely on expert elicitation	3	3		

Note. The frequency of each technique in the table is calculated as the number of studies employing that specific technique, such that studies utilising multiple techniques are counted for each relevant technique. The table summarises the techniques from 134 references. The remaining 57 references from the 191 selected for the SLR were not included as they did not use data analytics techniques (e.g. academic or industry review papers and theoretical articles) or there was insufficient information about the modelling techniques employed.

[a]The expert systems techniques used in the literature consist of expert elicitation, system dynamics and analytic hierarchy process.

are directed acyclic graphs that concisely represent "the probabilistic dependencies between a given set of random variables" (Nagarajan et al., 2013). Their ability to simultaneously reason multiple interrelated variables with complex dependence structures offers an interpretable *multi-risk* (or system) approach. Most studies employing Bayesian networks harness their diagnostic capability, inferring contributing factors and causal pathways of operational risk events in FS (e.g. Mittnik & Starobinskaya, 2010; Neil et al., 2009; Sanford & Moosa, 2015) and ENR (e.g. O'Shea et al., 2015; Pence et al., 2020). Researchers also use them for prediction, for example, of accidents in power plants (Groth et al., 2020), equipment faults in LNG production (Hassini & Zouairi, 2011) and cyber threats (Abidemi et al., 2014). While various structure (important variables and dependencies) and parameter (conditional probabilities) learning methodologies exist for Bayesian networks, most studies in the field rely on expert elicitation rather than raw data. This is time-consuming and manually demanding, and subjectivity and bias associated with experts' knowledge, experiences and perceptions pervade, compromising the objectivity and reproducibility of models. Nevertheless, easily incorporating domain knowledge and expert judgement is a unique advantage of the technique. 'Soft', qualitative aspects of an operating system, like social and organisational factors, can be captured and quantified, consistent with SoTeRiA (Sanford & Moosa, 2012). Including system factors that are typically less quantifiable yet inherent to operational risk is dominant in both FS (e.g. Sanford & Moosa, 2015) and ENR (e.g. Pence et al., 2019) data-driven ORM models, and reflects topical culture and conduct risks (Ocelewicz et al., 2021).

Although applied infrequently (4%), probabilistic risk assessments (PRA), encompassing fault and event trees, are used in ENR safety and reliability analysis for descriptive and diagnostic understanding (e.g. Guo et al., 2017). Through deductive- and inductive-logic, they represent faults leading to an undesired event and consequences following an initiating event as tree-like graphs (Mohaghegh et al., 2009). Traditional PRA methods inform controls to minimise system hazards. However, the linear interactions limit their effectiveness in analysing complex non-linear systems. Moreover, PRA is restricted to failures identified in past events, rather than considering influences in the broader operating system (Leveson, 2004).

Despite wide use in operational risk analytics, expert systems are not prevalent in the studies reviewed in this SLR (2%), with system dynamics (Forrester, 1961) and analytics hierarchy process (Saaty & Peniwati, 2008) most frequently applied. This primarily relates to the review's focus on more

Table 4. Summary of data inputs and variables in the literature.

Category	Data input / Variable	Frequency
Risk, Incident and Loss	Time and/or Date	21
	Consequence	
	Severity (loss amount, treatment cost, days lost, categorical rating)	19
	Type (monetary, injury, disability, death)	18
	Unstructured Textual Incident, Safety or Operational Reports[a]	15
	Cause	11
	Loss Frequency	11
	Event Logs[b]	9
	Audit / Investigation Reports	6
	Activity associated with Event	5
	Response / Mitigating Actions	4
	Risk Level	3
	Personnel Affected	2
Technical System	Geospatial Data	21
	Worksite Location	
	Personnel Location	
	Equipment Location	
	Traffic Density	
	Equipment Type	10
	Meteorological Conditions	10
	Information System Event Log	8
	Pressure	8
	Temperature	8
	Gas / Fuel Level	7
	Equipment Age	6
	Process Duration	6
	Transaction Log	6
	Geological Conditions	5
	Bearing Vibrations	4
	Hazardous Substance Involved	4
	Internet Traffic	4
	Compliance to Safety Practices	3
Organisational Structure	Role	10
	Business Line	5
	Firm Size	3
	Analytical Capability	2
	Industry	2
	Management Control	2
	Safety Policy, Training and Culture	2
	Financial Performance	1
Social / People	Age	5
	Gender	4
	Safe Behaviour	4
	Experience / Tenure	2
	Performance / Proficiency	2
	Qualification / Education	2
Environmental	Country	2
	Economy	2
	Commodity Market Price	2
	Exchange Rate	1
	Employment Rate	1
	Political Pressures	1
	Regulatory Environment	1
	Season	1

Note: The list of variables in the table is not exhaustive of those used in the literature, nor is it representative of the entire 191 publications included in the SLR. It is an overview of the most commonly employed variables across key studies.

[a]Unstructured textual reports may contain the information of the other risk, incident and loss variables in the table. However, these reports are considered as a separate data input as some studies use incident, safety and operational reports as their key input. The other loss variables are likely retrieved from structured event logs and databases.

[b]Event logs likely contain the information of other risk, incident and loss variables in the table, such as Time, Consequence, Cause, Response and Personnel Affected. However, event logs are listed as a separate data input as some studies do not specify the individual variables or use them as the key data input.

automated and objective analytics approaches, driven by raw operational data. By contrast, expert systems methodologies often rely on expert ratings, or if primary or secondary data sources are used, experts play a key role in development.

Hybrid approaches are also used, sometimes spanning multiple model families. They allow the technique most suitable for each analytics objective within the overall task to be used, improving predictive accuracy (e.g. Ahmed & Abraham, 2015). Combining traditional judgement-based approaches (PRA and expert systems) with more advanced mathematical models (e.g. Bayesian networks) enhances the application of each method. Pence et al.'s (2019) integrated-PRA framework (refined in Pence et al. (2020)) integrates text mining for preprocessing, Bayesian networks, fault and event trees and simulation to quantify the influence of organisational attributes on risk scenarios and determine critical root causes of failure.

3.2.3. Critical analysis of data

A wide variety of data sources are used with variables relating to five categories (Figure 3) – risk, incident and loss data; technical system factors; organisational structure factors; social or people factors; and environmental factors. These categories correspond to key risk theories (Rasmussen, 1997; Reason, 1990) and data collected to comply with regulation or standards (e.g. ISO 14224:2016). Table 4 summarises the most frequent variables.

Loss data are collected and used to estimate banks' operational risk regulatory capital, as per the Basel accords (2006). This historical data on operational losses and near-misses includes timing, loss amount, event type, cause and consequence descriptions. Similar information is collected for incidents in ENR organisations, along with the associated business activity, mitigating actions and personnel affected (e.g. Güven et al., 2016; Silva & Jacinto, 2012). Loss data typically constitute the dependent variable, providing past events to learn from. Events are recorded in structured databases, such as governance, risk and compliance or enterprise resource planning systems, or unstructured textual incident or operational reports.

Technical system factors measure or explain the technical features or operating conditions of system components, including equipment, information systems and the physical environment. Such factors are predominantly used in ENR settings. Often a scientific fact or operational link relates the variable to an incident, motivating its inclusion. Additionally, socio-technical systems theory (Rasmussen, 1997) reinforces incorporating technical system aspects into risk analysis. ENR equipment (e.g. power plants (Groth & Bensi, 2018)) are commonly monitored through metrics like pressure, temperature and bearing vibrations, as well as maintenance records. Geospatial, geological or meteorological data (e.g. Middleton & Sabeur, 2011; Wu et al., 2011) capture the physical environment. Research in FS considers technical factors like system access and traffic (e.g. Nagashree et al., 2018; Urabe et al., 2011), database queries (e.g. Fedushko et al., 2020) and transactions (e.g. Liu & Liu, 2016).

Organisational structure characteristics encapsulate management, governance structures and business procedures. Structural and cultural organisational aspects, including policies and the tone from executives, influence risk culture and thus incident occurrences, rationalising consideration of these factors (Mohaghegh et al., 2009). Personnel role and business line most frequently capture organisational structure across both FS and ENR applications (e.g. Sanchez-Pi et al., 2014). Variables gauging overall organisational performance are also considered, including financials and analytical capability (e.g. Embrechts et al., 2018). Some studies investigate the extent and effectiveness of management control and communication (e.g. Leveson, 2004; Moura et al., 2017). Organisational structure factors are commonly surveyed from experts, given their largely subjective and qualitative nature.

As per socio-technical systems theory (Rasmussen, 1997), *social and people* aspects of organisations and risk scenarios are important considerations. Understanding the interactions between people and technical processes within complex systems can inform system breakdowns and accidents (Rasmussen, 1990). Various variables can proxy human behaviour. Generic attributes, like age and gender, are used to proxy personality, responsibility and experience (e.g. Neil et al., 2005; Persona et al., 2006). Other research includes measures of skill – tenure, qualification and performance metrics (e.g. Amoako et al., 2020; Moura et al., 2017). In ENR-focussed studies, compliance with safety policy and training measure safe workplace behaviour (e.g. Moura et al., 2017; Xiaoyun & Danyue, 2010).

Environmental variables aim to incorporate factors external to the organisation that may influence the operational environment and thus risks. Studies consider country, macroeconomic indicators, political uncertainties and regulatory pressures (e.g. Eckle & Burgherr, 2013; Peters et al., 2009). Research only incorporates these minimally, rather focussing on the aforementioned internal factors organisations can control.

No study considers variables from all five data input categories simultaneously (Figure 6). Relatively few studies (7.1%) incorporate variables from a combination of four categories. Approximately 70% use only loss data, technical system factors or both. Therefore, there is scope to expand the factors included in a study, particularly given evidence of complex interrelationships between technical, social and organisational factors (e.g. Onoda et al., 2009). Furthermore, studies using only loss event data, without considering the factors and conditions for non-events (normal operating conditions), provide imbalanced views of the operating environment and hence questionable conclusions.

Data are generally collected from individual organisations; national or regulatory authorities, industry associations or consortia, that pool data from participating organisations across an industry or country; or public data sources, such as news providers, social media and financial reports. Subject matter experts are a highly utilised source of information, identifying risks and causal factors, defining dependencies and parameterising probabilities, particularly for Bayesian networks, PRA and expert systems. In such cases,

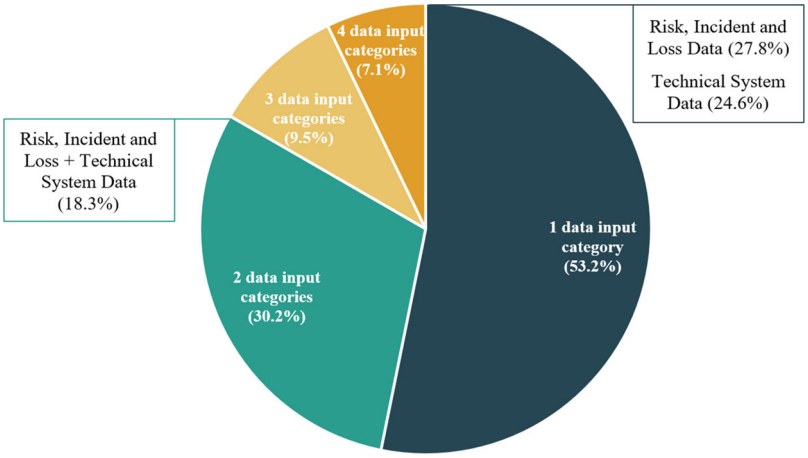

Figure 6. Frequency of the number of data input categories used simultaneously in research.
Note: Figure 6 reports the percentage of studies that use data from one, two, three or four categories of input data, with the three most highly utilised combinations of data input categories called out.

research succumbs to the variability of experts' experiences and inherent human biases in risk assessment. For example, experts may not have experienced many operational loss events, nor the range of business processes. However, internal loss data only capture the sample of events reported historically, overlooking those not reported. Some studies use synthetically generated data, particularly for proof of concept (Mazumder et al., 2021). Study type and risk perspective also influence the data source to an extent. For example, *empirical-micro* studies typically use data from an organisation, whereas *generic-macro* studies often use industry data.

4. Research themes

Review of bibliometric and content analyses (Section 3) revealed five core themes, segmenting how data analytics is used for different purposes in managing operational risks across FS and ENR organisations. The themes are:

(i) risk identification (11%)[6];
(ii) causal factors (25%);
(iii) risk quantification (17%);
(iv) risk prediction (21%); and
(v) risk decision-making (6%).

The prevalence of these themes differs for FS and ENR. Within each theme, there are several use cases or perspectives researchers take in achieving the theme's overall objective, which are presented as sub-themes. Each sub-theme is generally characterised by studies of a particular type, applying particular risk perspectives, analytics objectives, techniques and data inputs. To link the themes with the classification framework (Figure 3), (a) the following discussion and Tables 5–9 present the typical composition of studies as per the classification framework and sector(s) of focus in the literature; and (b) shading the breakdown of the literature as per the classification framework (Figure 4) highlights the portion of studies pertaining to each theme. Concluding each of the next sub-sections are critical analyses and future directions.

4.1. Theme I: risk identification

A group of FS and ENR *empirical-micro* studies and *generic-macro* frameworks leverage past public or organisational data to describe and identify key risks. These serve as horizon-scanning tools for risk managers by efficiently synthesising huge volumes of information – substantially more than humanly possible in manual document-based reviews – to objectively identify cases requiring attention. Used in both FS and ENR, NLP over textual data sources (e.g. news articles (e.g. Chu et al., 2020), financial statements (e.g. Wei et al., 2019) or risk and audit reports (e.g. Arumugam et al., 2016)) automatically distils current and emerging risks into a taxonomy. Arumugam et al. (2016) showcase this in their *empirical-micro* study, performing descriptive analytics with k-means clustering on risk phrases extracted from reports of offset wells using NLP to streamline well drilling planning and execution.

A difference for FS is that, further to text mining risks, studies on banks and insurers have also identified risks through anomaly detection and conformance checking. Anomaly detection mechanisms using unsupervised techniques on internal records alert risk managers to suspicious transactions (e.g. Palshikar & Apte, 2013) or customers (e.g. Xiaoyun & Danyue, 2010). Process mining, comparing historic workflow logs with process blueprints, highlights non-conformances (e.g. Huang et al., 2012). Table 5 presents more detail.

4.2. Theme II: causal factors

A quarter of the literature aims to diagnose the factors contributing to operational loss events. Although applied to information security vulnerabilities in financial institutions in a few studies (e.g. Lien, 2012), causal factors research mostly relates to safety and equipment failure incidents within ENR (e.g. Milana et al., 2019; Onoda et al., 2009). This application is revolutionising manual root cause and failure pathways analyses traditionally completed in safety-critical industries in pre-project planning, post-incident reviews or general risk management functions at infrequent intervals. Applying diagnostic analytics across multiple disparate data streams efficiently provides risk managers with more scientific and detailed understandings of the factors leading to incidents. Ultimately, they inform policy, training, maintenance and other operational decisions, yielding safe and successful operations (Stojadinovic et al., 2012). For example, Leite's (2019) *empirical-micro-diagnostic* analysis using logistic regression on hydropower plants' technical data (system monitoring, maintenance and performance) identifies factors that significantly indicate longer shutdown periods – factors helpful to operators in prioritising inspections and reducing system downtime.

Causal findings are often very context-specific as most studies are *empirical-micro* investigations, applying statistical and machine learning techniques to loss and technical systems data for specific incidents in ENR or FS. Another example is Dong et al. (2019) using k-means clustering and simulation to assess the factors that heighten the likelihood of residential transformers overloading in electricity distribution. Table 6 presents more detail.

Alternate to this *micro* risk perspective and focus on technical system factors in ENR and FS, a portion of the causal factors literature is distinct to ENR – developing generic *multi-risk* frameworks for site safety, underpinned by systems theory (Reason, 1990) and SoTeRiA (Rasmussen, 1997). These Bayesian network or hybrid models (e.g. Pence et al., 2020; Rodriguez-Ulloa, 2018) show that human errors do not occur randomly but are linked to system factors with complex interrelationships. This highlights the value of incorporating multiple risks and factors (technical system, social, organisational or environmental) into a single model. A noteworthy contribution is the spatio-temporal simulation module within Pence et al.'s (2019) integrated-PRA framework, which seeks to progress traditional, static PRA to dynamic analysis, considering temporal and spatial dimensions. There is opportunity to advance data-driven causal factors analyses to be live, continuous and consider dependencies across a large range and volume of data sources. While work is needed to fully operationalise this, currently ENR is more advanced than FS. Causal analysis approaches applied in ENR contexts should be investigated for FS to gain insight into the underlying factors driving loss events, rather than simply quantifying risk levels.

4.3. Theme III: risk quantification

Some studies quantify operational risk to assess inherent and residual risk levels and compare them with risk appetite. In both FS and ENR, quantification methodologies broadly involve combining estimates of the frequency and severity of events. Several *empirical-* and *generic-micro* studies apply descriptive analytics with PRA or Bayesian theory to calculate the value of potential losses throughout the oil and gas supply chain (e.g. Eckle & Burgherr, 2013). In Hamedifar et al. (2015), empirically quantifying the frequency, consequences and thus total risk of loss of containment from LNG carriers using PRA on marine traffic, ship operational and environmental data enables quantitative analysis of cost-effective risk reduction strategies. Risk quantification, however, is most commonly applied to estimate capital in the banking sector, strongly motivated by regulation (e.g. Dutta & Perry, 2006).

Specific to FS, capital estimation methodologies, driven by traditional LDA or Bayesian networks, mostly take *macro* perspectives (Table 7), aligning to the Basel II standards (e.g. Chavez-Demoulin et al., 2016; Mittnik & Starobinskaya, 2010). These aggregate approaches lack risk sensitivity by not explicitly considering individual risks, nor related operating factors. Therefore, capital estimations are not as sensitive to fluctuations in the profile of different operational risk events, nor the state of the underlying risk drivers. Yet even studies that only consider dependencies between aggregated groups (operational risk types and business lines) find avalanche-like effects occur from high levels of dependencies (Mittnik & Starobinskaya, 2010). Adopting a *multi-risk* perspective using Bayesian networks, and incorporating the causal factors, provides greater transparency of such effects (e.g. Sanford & Moosa, 2015). It is more holistic and proactive as it models "uncertainty about the process that generates losses as well as the distribution of losses that might result" (Neil et al., 2005, p. 971). This allows financial institutions and regulators to understand what is driving capital levels, and thus ways to better mitigate risks, in turn reducing capital requirements.

While the capital regulations driving risk quantification are unique to FS, ENR could consider the FS capital estimation approaches that measure the

Table 5. Research perspectives in theme I: risk identification.

Theme I: risk identification						
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data Inputs
Text Mining Risks [FS, ENR]	• Identify risks and construct risk taxonomies and registers for organisations and industries by using text mining techniques • Automatic text mining methods found effective and efficient at identifying current and emerging risks • Textual data sources may present biases towards high severity, low frequency incidents; publicly disclosed information; or organisations with high press coverage	Empirical Investigation	Micro Macro	Descriptive	Natural Language Processing	• News Articles (Chu et al., 2020; Leidner & Schilder, 2010; Nugent & Leidner, 2017) • Financial Statements (L. Wei et al., 2019) • Internal Risk and Audit Reports (Arumugam et al., 2016; Satoh & Samejima, 2019)
Anomaly Detection [FS]	• Identify outliers and monitor deviations from normal operating conditions, thus highlighting higher risk cases (e.g. customers or transactions)	Generic Framework	Micro	Descriptive	Clustering	• Customer Records (Xiaoyun & Danyue, 2010) • Transaction Records and Suspected Money Laundering Case Reports (Palshikar & Apte, 2013)
Auditing and Conformance Checking [FS]	• Identify non-conformances of information systems and processes and highlight areas for process improvement • Greater depth of audit analyses and less resource-intensive and invasive for personnel than traditional audits	Generic Framework	Micro	Descriptive	Process Mining (Caron et al., 2012, 2013; Huang et al., 2012)	• Actual Workflows from Information System Event Logs • Business Process Blueprints and Rules

likelihood and consequences of operational risks on a continuous scale to yield more precise indicators of their changing risk profile. The *multi-risk* quantification techniques may be particularly relevant in the large-scale and complex socio-technical operating environments of ENR organisations, essentially offering more precise quantification of causal factor analyses.

4.4. Theme IV: risk prediction

Over one fifth of the literature predicts aspects of *micro* operational risks, including the type of risk events imminent, the timing, probability or severity. As an indicative example, Yin et al.'s (2021) *empirical-micro* investigation predicts gas kick events in deepwater drilling earlier than traditional methods using a long short-term memory ANN on sensor data. Risk prediction is most prominent in ENR, with studies predicting safety hazards, accidents, equipment failures and maintenance requirements (e.g. Li et al., 2020; Xie et al., 2019). Across FS, information security risks and fraud detection are the foci (e.g. Abidemi et al., 2014; Jans et al., 2010b). Regardless of industry, risk prediction is considered from two subtly different perspectives (Table 8); yet all empirically investigate various supervised statistical and machine learning techniques or Bayesian networks. No technique is definitively the best in either sector, attesting to the application-specific nature of studies and lack of common evaluation metrics.

Predictive risk algorithms are particularly valuable when implemented at the first line of defence[7], meaning personnel "on-the-ground" get access or alerted to context-sensitive risk information to efficiently inform decision-making and prevent incidents. Sanchez-Pi et al. (2015) exemplify this, developing an early warning detection system for offshore oil extraction and processing. Powered by association rule mining across spatial, temporal and technical data, it predicts risky scenarios given employees' locations and organisational roles and notifies them in real-time.

Most risk prediction applications relate to incidents observed relatively frequently, learning from past known occurrences. A smaller body of research aims to predict rare or even unseen events by tracking factors in the system over time (Urabe et al., 2011) or combining expert knowledge with machine learning (Milkau & Bott, 2018). However, further innovations are required to model unknowns, given complexities surrounding data limitations. The extreme infrequency of unknown unknowns means a much larger sample of data is needed in order to model these events with existing techniques, and yet

Table 6. Research perspectives in theme II: causal factors.

	Theme II: causal factors					
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data inputs
Safety Incident Causality Research [ENR]	• Analyse the causality of accidents and near-miss safety incidents in coal and oil mine sites • Successfully identify common factors contributing to safety incidents • Lack of structured data sources in safety-critical industries somewhat limits the potential analyses and generalisability of results	Empirical Investigation	Micro	Diagnostic	• Statistical Frequency Analysis (Stojadinovic et al., 2012) • Multivariate and Logistic Regression (Amoako et al., 2020; Leite, 2019) • Decision Trees (Hajakbari & Minaei-Bidgoli, 2014; Persona et al., 2006) • Association Rule Mining (Parhizi et al., 2009; Y. Wu et al., 2015) • Clustering (Milana et al., 2019; Parhizi et al., 2009) • Natural Language Processing (Bouveret, 2019; Y. Wang et al., 2018) • Pre-processing (Milana et al., 2019)	• Unstructured Textual Incident, Safety or Operational Reports
Equipment Failure Modes Analysis [ENR]	• Identify warning signs of troubling conditions or failure of equipment in asset-intensive organisations (e.g. oil and gas extraction (Johnston & Guichard, 2015), hydroelectric power production (Onoda et al., 2006, 2009), electricity distribution (Dong et al., 2019)) • Unknown, and some unexpected, patterns and correlations uncovered, informing maintenance requirements and successful operations	Empirical Investigation	Micro	Diagnostic	• Support Vector Machine • Scenario Analysis and Simulations	• Technical Readings from Equipment Sensors
Information Security Vulnerability Detection [FS]	• Identify employee attributes and behaviours that threaten the security of customer data and knowledge within organisations • More comprehensive and efficient than existing manual audits of randomly sampled data logs	Empirical Investigation	Micro	Diagnostic	• Decision Trees • Fuzzy Decision Trees	• Database Query Logs (Lee et al., 2013; Lien, 2012; Lien et al., 2011)
Qualitative Systems Approach Root Cause Analysis (Leveson, 2004; Pika et al., 2013) [ENR]	• Structured, qualitative frameworks based on systems, or human error, theory to retrospectively analyse and identify the factors causing an accident • Assist in engineering safer systems from technical, managerial, organisational and regulatory perspectives	Generic Framework	Micro	Diagnostic	• System Dynamics • Influence Diagrams • Fault Tree Analysis • Event Tree Analysis	• SME • Incident Reports

(*continued*)

Table 6. Continued.

	Theme II: causal factors					
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data inputs
	• Require subject matter experts to review comprehensive accident reports and, fundamentally, backward-looking					
Socio-Technical Risk Analysis (SoTeRiA) extended to Organisational Factors (Pence et al., 2014; 2019; 2020) [ENR]	• Extension of SoTeRiA to identify human, system and organisational factors that influence human errors • Approach "guides 'data analytics' with 'theory'" (Pence et al., 2019, p. 242) to leverage the value of data while avoiding incomplete or inaccurate results from solely data-driven approaches • Incorporates spatio-temporal to progress traditional, static probabilistic risk assessment to dynamic analysis	Generic Framework	Multi-risk	Diagnostic	Hybrid Models of: • Bayesian Networks • Fault Tree Analysis • Event Tree Analysis • Event Sequence Diagrams • System Dynamics • Text Mining • Simulation	Incident Reports, SME, Industry or Operational Data on: • Technical System Factors • Social Factors • Organisational Factors
Bayesian Network Operational Risk System Modelling (Neil et al., 2009; Qazi et al., 2015; Rodriguez-Ulloa, 2018; Sanford & Moosa, 2015) [FS, ENR]	• Model multiple operational risks and their causal factors using Bayesian Networks	Generic Framework	Multi-risk	Diagnostic	• Bayesian Networks	• SME • Loss Event Data
Other Frameworks for Micro Risk Causal Factors Analysis [FS, ENR]	• Identify causal factors of specific risk events in specific industry contexts	Generic Framework	Micro	Diagnostic	• Association Rule Mining (He & Song, 2009; Kenett & Salini, 2010; Xiaorong et al., 2009) • Sequential Pattern Mining (Cao et al., 2016) • Artificial Neural Network (Setiono et al., 2006) • Self-organising Maps (Moura et al., 2017)	• Loss Event Data • Incident Reports • Technical System Data (e.g. Inventory and Transaction Records)

Table 7. Research perspectives in theme III: risk quantification.

Theme III: risk quantification						
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data inputs
Traditional LDA Capital Estimation [FS]	• Estimate regulatory capital required using the LDA, with various techniques used to parameterise the frequency and severity distributions and the dependency structures • Lack risk sensitivity, such that do not consider possible interconnections between individual operational risk events, nor with other factors, which may offer additional insight about an operational environment	Empirical Investigation Generic Framework	Macro	Predictive	• Parametric Distribution Fitting with Copula Functions (Dutta & Perry, 2006; Hao, 2013; Yinghui Wang et al., 2017; Zhu et al., 2014) • Semi-parametric Extreme Value Theory (Chavez-Demoulin et al., 2016; Dutta & Perry, 2006; Embrechts et al., 2018) • Non-parametric Empirical Sampling (Dutta & Perry, 2006)	• Internal Loss Data (including loss amount, relevant business line, risk type and year of loss) • Business Environment and Internal Control Factors
Macro Capital Estimation using Bayesian Networks (Azar & Mostafaee Dolatabad, 2019; Lambrigger et al., 2007; Mittnik & Starobinskaya, 2010) [FS]	• Estimate regulatory capital required using a Bayesian network structure with nodes resembling the elements of the LDA (frequency, severity and total loss distributions) and AMA (risk categories and business lines) • Bayesian Networks inherently incorporate dependencies between business lines and risk types • Avalanche-like effects on expected losses occur from high levels of dependencies • Directed acyclic graphical structure allows for the representation of idiosyncratic risk factors for each business line, risk factors common to several business lines and risk contagion between business lines • Lack risk sensitivity • Heavy reliance on subjective inputs from experts, thus inherent human bias and influenced by experts' experiences and perceptions	Generic Framework	Macro	Diagnostic	• Bayesian Networks	• SME • Loss Event Data
Multi-risk Capital Estimation using Bayesian Networks (Neil et al., 2005; 2009; Sanford & Moosa, 2012, 2015) [FS]	• Estimate regulatory capital required by modelling the causal factors relating to operational risk events as random variables in a Bayesian Network • Realistic and reproducible total loss estimates that comply with the AMA requirements, and also gain understanding of the underlying causal factors of events • Only select few operational risk events incorporated, along with small sample of causal factors, despite adopting a systems approach • Heavy reliance on subjective inputs from experts	Generic Framework	Multi-risk	Diagnostic	• Bayesian Networks	Experts and Loss Event Data on: • Operational Procedures • Staff Skills • Working Environment
Micro Risk Quantification [ENR]	• Quantify the level of risk of specific risk events in specific industry contexts	Empirical Investigation Generic Framework	Micro Multi-risk	Descriptive	• Parametric Distribution Fitting (Harmantzis & Malek, 2004) • Bayesian Theory (Eckle & Burgherr, 2013) • Probabilistic Risk Assessment (Hamedifar et al., 2015) • Fuzzy Theory (Tubis et al., 2020)	• Loss Event Data • Technical System Data • Environmental Data • Experts

Table 8. Research perspectives in theme IV: risk prediction.

Theme IV: risk prediction						
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data inputs
Prediction by Monitoring Causal Factors [ENR]	• Monitor conditions of critical causal factors identified from diagnostic analysis as a form of early warning risk detection, often embedded in notification systems	Empirical Investigation	Micro Multi-risk	Diagnostic Predictive	• Association Rule Mining (Fang et al., 2017; Nagashree et al., 2018; Sanchez-Pi et al., 2014, 2015) • Anomaly Detection (Haustein et al., 2008)	• Loss Event Data • Technical System Data
Single Risk Prediction Model [ENR]	• Identify causal factors and predict risk events using a single technique	Empirical Investigation	Micro	Predictive	• Bayesian Networks (Abidemi et al., 2014; Groth et al., 2020; Hassini & Zouairi, 2011) • Association Rule Mining (Y. C. Wei et al., 2018) • Decision Trees (Zhang et al., 2017) • Random Forests (Liang et al., 2020; Zhang et al., 2017) • Analytic Hierarchy Process (Wu et al., 2011) • Artificial Neural Network (Yin et al., 2021) • Hybrid Models (Ahmed & Abraham, 2015; Guo et al., 2017)	• Loss Event Data • Technical System Data • Environmental Data • Experts

a lack of data remains on these cases; hence there are still open questions.

4.5. Theme V: risk decision-making

All studies in the field intend to inform risk decision-making to some extent, but a small number spanning both FS and ENR aim to primarily prescribe the optimal decision. This is approached from two perspectives (Table 9). First, automated risk decision-making systems are explored conceptually, calculating the optimal mitigation strategy and resource allocation based on reward maximisation techniques (El Bouchti et al., 2017; Rahmes et al., 2013). For example, Rahmes et al.'s (2013) *generic-macro* framework applies prescriptive analytics through game theory analysis (combining probabilistic event predictions from decision trees and event sentiments derived from NLP on historical records of reactions) to populate reward matrices for the most efficient and effective allocation of resources. Second, more holistic, qualitative frameworks for applying data analytics to ORM that ultimately inform risk decision-making are developed (e.g. Groth & Bensi, 2018), consisting of a similar structure – (i) data management architecture integrating disparate data sources, (ii) a model-building module aligned to the risk management objective and (iii) interpretation and visualisation of results. Progression towards autonomous risk decision-making systems is in its infancy for FS and ENR. As organisations in these sectors navigate digital transformation, while maintaining compliance in increasingly regulated environments, it is critical future methodologies are transparent to ensure risk managers, executives and regulators have visibility over the workings, bolstering confidence in the approach.

5. Discussion and conclusions

Analysing the literature applying data analytics to ORM across the FS and ENR sectors has highlighted several key distinctions and similarities in their approaches. The following summarise these by the components of the literature classification framework.

Risk perspective: Overall, there is a high concentration of empirical and *micro* studies, limiting the generalisability of findings to other contexts, although methodologies may be transferrable. This is particularly evident among the ENR literature, with most studies either investigating the causal factors of or predicting specific safety incidents or equipment faults. In comparison, a larger proportion of FS research employs a *macro* perspective, analysing operational risk as a holistic category. Although, some FS operational risks, including fraud and information security threats, lend themselves to *micro* diagnostic or predictive analyses. Across both sectors, research from a *multi-risk* perspective remains largely theoretical and qualitative, with the instances of quantitative approaches relying on subjective data elicited by experts (e.g. Pence et al., 2020; Sanford & Moosa, 2015). Yet the COVID-19 pandemic showcases how a single event

Table 9. Research perspectives in theme V: risk decision-making.

Theme V: risk decision-making						
Research perspective [Focus sector]	Description and findings	Study type	Risk perspective	Analytics objective	Analytics techniques	Data inputs
Automated Risk Decision-making Systems [FS, ENR]	• Prescribe the optimal action or strategy, given knowledge of the past, current and predicted future states • Few quantitative case studies testing the application of these theoretical models	Generic Framework	Micro Macro	Prescriptive	Reward Maximisation Methods • Deep Reinforcement Learning (El Bouchti et al., 2017) • Game Theory (Rahmes et al., 2013)	• Loss Event Data • Experts
Data-driven ORM Framework (Groth & Bensi, 2018; Johnson, 2010; Middleton & Sabeur, 2011; Wu, 2020) [FS, ENR]	• Holistic, qualitative frameworks for applying data analytics to manage a specific operational risk to ultimately inform risk decision-making, with the general structure of (i) a data management architecture to integrate disparate data sources, (ii) a model-building module aligned to the risk management objective and (iii) interpretation and visualisation of results • Specified for specific risk contexts and mostly tested on only one case study, limiting generalisability • Do not consider multiple risk events simultaneously • Require considerable manual set up and intervention by experts with risk and data analytics knowledge	Generic Framework	Micro	Descriptive Diagnostic Predictive Prescriptive	• Classification • Prediction • Time-series Analysis • Association Analysis • Clustering • Anomaly Detection	• Loss Event Data • Technical System Data • Experts

can greatly affect many dimensions of an organisation's operational risk profile (e.g. employee and customer health, home workplace safety, information security in the remote work environment, supply chain (Actuaries Institute, 2020; Evans, 2020)) and the wider economy (McKibbin & Fernando, 2021). This systemic disruption and risk contagion foregrounds the importance of *multi-risk* perspectives in all sectors. Hence, the authors recommend *multi-risk* perspectives persist in future research and organisational practice.

Analytics objective: Much of the FS and ENR research extends beyond backward-looking descriptive investigations, which include text mining risks to form taxonomies or detecting anomalies and non-conformances in financial institutions' systems. ENR studies are predominantly diagnostic and predictive analyses of accidents involving personnel or systems. Despite the dominance of diagnostic studies determining *what* factors cause risk events, few approaches attempt to understand *how* causal factors influence an operational risk profile. Such analysis could provide risk managers with more precise insights to inform resource allocation for risk control, mitigation and monitoring. In contrast, FS research is more advanced in quantifying uncertainty and modelling dependencies between operational risks across an organisation, with these studies pioneering Bayesian networks for ORM (Neil et al., 2005) and employing copula functions in traditional statistical approaches (e.g. Peters et al., 2009). These learnings from FS will be important as ENR research progresses to understanding how risk profiles change at more granular levels. Research into prescriptive analytics for risk decision-making is immature across both FS and ENR, noting some conceptual developments, yet further research is required to operationalise such systems. This is not surprising given fully autonomous systems are currently too *black box* for most organisations and regulators.

Analytics techniques: Reflective of the analytics objectives of focus for each sector, traditional statistical techniques are mostly used in FS research (as per capital regulation), whereas supervised statistical and machine learning techniques are prevalent for ENR. Additionally, relatively rudimentary methodologies, which lack discriminative and predictive power, such as association rule mining and fault and event tree analyses, are frequently employed for safety and reliability analysis in ENR contexts. Extending these tree-like PRA methodologies, Bayesian networks are increasingly being used to model operational uncertainties across both sectors. Although a range of modelling techniques have been investigated, few studies compare techniques (e.g. Monish & Pandey, 2020; Yin et al., 2021). Understanding the empirical and theoretical appropriateness of techniques across various contexts is pivotal in attaining the most effective, efficient and generalisable approach. Within studies, limited emphasis is placed on model performance assessment. Yet this is pertinent, particularly given heightening awareness and attention among practitioners

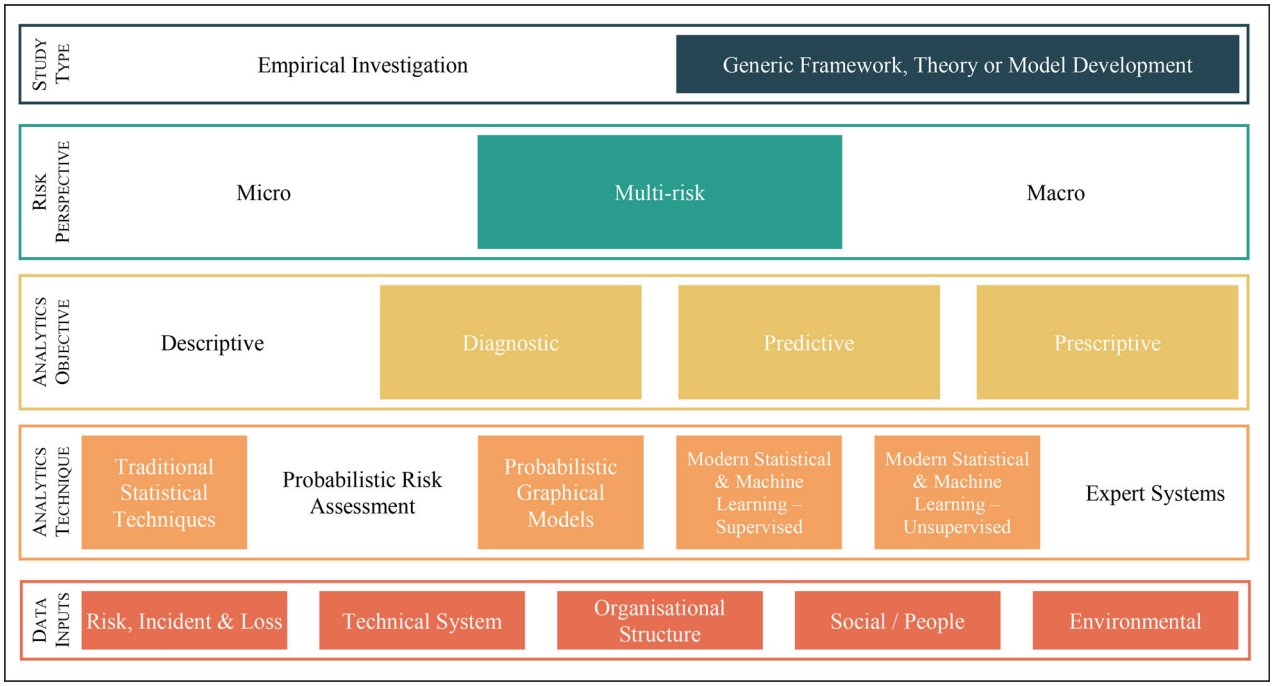

Figure 7. Main areas for future research as per literature classification framework.
Note: The shaded elements of the framework highlight the current gaps in the research area. Various combinations of the shaded components across the five layers would benefit from future research.

and regulators on model ethics and emerging technology risks. Researchers should consider evaluation more along the lines of Dutta and Perry (2006) and Xie et al. (2019).

Data inputs: Researchers have used a wide variety of data and variables with loss data forming the basis of the majority of studies, sourced from either structured risk databases (typical of FS) or unstructured textual reports (common in ENR). FS applications largely rely on loss data and expert input, whereas research using technical system, social and organisational data is more prevalent in ENR. Leveraging raw, continuously updated data in organisations as in much of the existing ENR research will be critical in the data-driven ORM evolution, rather than relying on static, subjective data elicited by experts. Only a small selection of factors relating to technical system, organisational structure, social and people or macro-environmental aspects are considered simultaneously. If a wider set of factors is considered (similar to Persona et al. (2006) and Pence et al. (2014) who consider incident consequences, timing, technical system information, organisational roles and peoples' demographic characteristics), additional insights about complex interconnections and leading indicators may be uncovered. Extending the sample period of factors to normal operating conditions, rather than only at loss event times (i.e. in the spirit of condition monitoring in ENR (e.g. Marquez et al., 2020; Onoda et al., 2006)), could provide a more holistic and balanced representation of the operating environment, and in turn more precise risk insights.

A final observation: Apart from the capital estimation literature in banking, much of the research in FS and particularly ENR is framed as a classification problem, with a binary (e.g. "risk" or "no risk"), multinomial (e.g. accident type) or ordinal (e.g. "high", "moderate" or "low" risk) response. These discrete risk views linger from traditional ORM practices – for example, likelihood-severity risk matrices with qualitative scales lead to bucketed risk ratings from which several risks with the same rating cannot be differentiated (Ashley, 2020). Operational risk lies, however, on a continuum of both time and magnitude. Hence, quantifying it as a continuous probability would provide more realistic and detailed representations of dynamic operations.

This SLR's comparison between FS and ENR highlights opportunities for each sector to learn from approaches applied by the other. Similarly, common limitations and underdeveloped areas across the sectors inform avenues for future research. Figure 7 depicts the main research gaps on the classification framework, and the following outline recommendations for future research across the sectors reviewed.

- Building on the banking industry's approach to quantifying operational risk as a continuous probability (e.g. Dutta & Perry, 2006), and combining this perspective with ENR's detailed causal factors analysis (e.g. Moura et al., 2017; Pence et al., 2020), further research is needed to better understand how causal factors alter the probability of an operational loss event.

Table 10. Summary of research themes.

Research theme	Analytics objective	Analytics techniques
(i) Risk Identification	Descriptive	NLP, Clustering, Processing Mining
(ii) Causal Factors	Diagnostic	Association Rule Mining, Decision Tree, NLP, Bayesian Network
(iii) Risk Quantification	Predictive, Diagnostic	Traditional Statistical Distribution Fitting (LDA)
(iv) Risk Prediction	Predictive	ANN, Decision Tree, Association Rule Mining, Bayesian Network
(v) Risk Decision-making	Prescriptive	Deep Reinforcement Learning

- Most existing ORM solutions in FS and ENR are static, such that their results reflect the risk at a single time point. Pence et al. (2019), however, introduced temporal dimensions to PRA. A logical next step to fully benefit from data analytics is to develop dynamic learning systems that reliably and repeatedly detect changes in risk profiles. This extends to incorporating decision and utility nodes into predictive frameworks, as per El Bouchti et al.'s (2017) suggestion of reinforcement learning, to enable prescriptive analytics – live, actionable information on current and emerging risks. A model, possibly hybrid, with flexibility to adapt to environmental changes and incorporate new risks without extensive modification is of great value and warrants further research.
- Past operating conditions, and perhaps previous risk levels, may inform the future level of risk. Hence, research into techniques that allow for time-series relationships and dependence between risks and business processes, along the lines of Nwafor et al. (2019), is still needed in both sectors.
- Future research should investigate methods and inputs that reduce the reliance on subjective, expert-elicited data to avoid the requirement for manual human intervention, limit human bias and improve consistency. Methods may include greater usage of raw operational and risk data as in existing approaches within ENR (e.g. Yin et al., 2021); leveraging near-miss data to more objectively infer probability distributions; or considering ways to more reliably use expert input, such as applying credibility weightings based on the accuracy of expert estimates historically (similar to Sanford and Moosa (2015)), using experts to validate model predictions and incorporating this in a feedback loop, or other mechanisms from expert elicitation literature. In achieving this, researchers should be mindful of the limitations of data-driven approaches, surrounding inaccuracies when predicting risk events that have historically occurred at low frequencies (Milkau & Bott, 2018).
- In both sectors, studies comparing the performance of various techniques under a single operational context, like Monish and Pandey (2020), are necessary to provide clarity of their relative effectiveness. Establishing standardised metrics to evaluate the performance of data-driven ORM approaches and existing qualitative risk management strategies would enhance comparability.
- Given existing research's reliance on unstructured data, particularly in ENR, research extending existing work (e.g. ISO, 2016) should be conducted to define systems and procedures for more accurate, complete and efficient collection and storage of operational and risk data. This would better enable data-driven ORM solutions to be developed and updated in practice.

Despite the nuances in how data analytics is applied to ORM in FS and ENR, there are, however, many similarities, which are largely reflected in the core research themes. Indeed, the industry-agnostic nature of the themes makes them generalisable beyond FS and ENR[8] (e.g. Rajesh, 2020a; 2020b). They highlight to risk managers and researchers in the field the key avenues through which organisations could benefit from data analytics in ORM. Understanding this is critical given heightening regulations and standards surrounding effective ORM across various industries. Table 10 summarises the themes with respect to the analytics objectives and techniques commonly applied. Taking this industry-agnostic view, a myriad of research opportunities exist to enhance the value of data-driven ORM approaches to all organisations:

- The coverage of operational risks, operational contexts and explanatory covariates needs to be extended. This should culminate in an objective, holistic data-driven ORM framework that considers the interconnectedness of risks across all operational categories, as well as a broad set of organisational and external factors. As with the COSO and ISO ERM frameworks, such research should develop a tool that is transferable within and between industries. This research agenda is strongly supported by academia and industry (Azvine et al., 2007; Choi & Lambert, 2017).
- A core weakness across most of the literature is the lack of implementation to real organisational data. This is poignant given digital processes can be scaled at near zero marginal cost, unlike traditional manual document-based risk management processes – a key advantage of data-driven ORM.

Seeking such opportunities should be a focus of future research and will require cooperation with industry partners. To aid in this, research will need to translate conceptual models into computer models and develop valuable visualisations of results.

- When reflecting on the management of the COVID-19 pandemic, it will provide an opportune stress test period to historically assess the robustness and flexibility of data-driven ORM solutions developed.

This paper has thoroughly reviewed the role of data analytics within ORM across FS and ENR through a methodical process. A novel classification framework has conceptualised the area, highlighting key insights about the methodologies and data in prior research. Five core research themes have been presented, alongside future research recommendations to advance ORM within and beyond FS and ENR.

Notes

1. The largest aggregate operational loss event categories between 2014 and 2019, being Clients, Products & Business Practices for banks and Execution, Delivery & Process Management for insurers (Operational Riskdata eXchange Association, 2020), reflect this environment. As defined in the Basel II operational loss event type classification, losses relating to Clients, Products & Business Practices arise "from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product", and losses relating to Execution, Delivery & Process Management arise "from failed transaction processing or process management, from relations with trade counterparties and vendors" (Basel Committee on Banking Supervision, 2006).
2. Preliminary analysis on the full set of 2,538 references retrieved from the initial literature review search informed the high-level findings of the research applying data analytics to risk management generally presented in the introduction.
3. Supplementing Figure 4, a list of the references included in the SLR with their relevant research theme and classification framework assignment is available upon request.
4. SoTeRiA stems from Rasmussen's (1990) perspective that bad outcomes are the result of normal interactions between people and systems, rather than human errors or deficiencies in human reliability (i.e. there are structural problems in how humans operate in large and complex systems).
5. The Basel II measurement methodologies include the Basic Indicator Approach, the Standardised Approach and the Advanced Measurement Approach (AMA) (Basel Committee on Banking Supervision, 2006). In response to the GFC, Basel III introduced a new Standardised Measurement Approach (SMA), which will be implemented in January 2023 (Basel Committee on Banking Supervision, 2020), with the aim to remove evaluation discretion associated with the AMA (Basel Committee on Banking Supervision, 2010). The SMA also aggregates operational risks to an organisational level, using a standard risk capital charge formula with business size indicator and internal loss multiplier components.
6. These percentages of references in the SLR for each theme do not sum to 100% since not all references were methodology articles developing or implementing a specific model or framework, and thus were not categorised into one of the five research themes. The remaining 20% of references were academic or industry review papers or theoretical articles.
7. As per the Three Lines of Defence model – a best practice framework for the systematic delegation and coordination of risk management duties within an organisation – the first line of defence is operational management, whose function is to "own and manage risks" (The Institute of Internal Auditors, 2013).
8. The generalisability of the themes to other sectors is showcased in the original search results retrieved. Studies from other industries, including industrials (e.g. manufacturing and construction), aviation and utilities (e.g. transportation, telecommunications and water supply), were also prevalent and when reviewed, related back the five core themes.

Acknowledgements

The authors thank the anonymous reviewers for their considered and extensive feedback throughout the revisions process. It has been very valuable in enhancing the quality and contribution of this paper.

Funding

This work was supported by KPMG Australia and an Australian Government Research Training Program Scholarship.

ORCID

Nikki Cornwell http://orcid.org/0000-0002-6424-2303
Christopher Bilson http://orcid.org/0000-0001-7391-4397
Adrian Gepp http://orcid.org/0000-0003-1666-5501
Steven Stern http://orcid.org/0000-0003-4048-5121
Bruce J Vanstone http://orcid.org/0000-0002-3977-2468

References

Abbass, W., Baina, A., & Bellafkih, M. (2020). Evaluation of security risks using Apriori algorithm [Paper presentation]. 13th International Conference on Intelligent Systems: Theories and Applications, SITA 2020.

Abidemi, B. M., Kayode, A. B., Favour-Bethy, T. A., & Otasowie, I. (2014). A bayesian network model for risk management in cyber situation [Paper presentation]. World Congress on Engineering and Computer Science 2014, WCECS 2014.

Actuaries Institute. (2020). *Pandemic briefing: Addressing COVID-19 impacts on non-financial risks in FCRs.*

https://actuaries.logicaldoc.cloud/download-ticket?ticketId=9c78fa88-27f4-444e-ba00-5762b561952e

Ahmed, N., & Abraham, A. (2015). Modeling cloud computing risk assessment using ensemble methods [Paper presentation]. 4th World Congress on Information and Communication Technologies, WICT 2014.

Akinosho, T. D., Oyedele, L. O., Bilal, M., Ajayi, A. O., Delgado, M. D., Akinade, O. O., & Ahmed, A. A. (2020). Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, *32*, 101827. https://doi.org/10.1016/j.jobe.2020.101827

Akter, S., Michael, K., Uddin, M. R., McCarthy, G., & Rahman, M. (2022). Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. *Annals of Operations Research*, *308*(1–2), 7–39. https://doi.org/10.1007/s10479-020-03620-w

Amoako, R., Buaba, J., & Brickey, A. (2020). Identifying risk factors from MSHA accidents and injury data using logistic regression. *Mining, Metallurgy and Exploration*, *38*(1), 1–19.

Araz, O. M., Choi, T. M., Olson, D. L., & Salman, F. S. (2020). Role of analytics for operational risk management in the era of big data. *Decision Sciences*, *51*(6), 1320–1346. https://doi.org/10.1111/deci.12451

Aria, M., & Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arumugam, S., Gupta, S., Patra, B., Rajan, S., & Agarwal, S. (2016). Revealing patterns within the drilling reports using text mining techniques for efficient knowledge management [Paper presentation]. 2016 SPE Eastern Regional Meeting, ERM 2016. https://doi.org/10.2118/184062-MS

Ashley, M. (2020). *Risk heat maps have failed us… now what?* https://www.linkedin.com/pulse/risk-heat-maps-have-failed-us-now-what-mike-ashley/

Australian Business Deans Council. (2019). 2019 Australian Business Deans Council (ABDC) *journal quality list.*

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, *253*(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023

Aven, T., & Flage, R. (2020). Foundational challenges for advancing the field and discipline of risk analysis. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *40*(S1), 2128–2136.

Azar, A., & Mostafaee Dolatabad, K. (2019). A method for modelling operational risk with fuzzy cognitive maps and Bayesian belief networks. *Expert Systems with Applications*, *115*, 607–617. https://doi.org/10.1016/j.eswa.2018.08.043

Aziz, S., & Dowling, M. (2019). Machine learning and AI for risk management. In T. Lynn, J. G. Mooney, P. Rosati, & M. Cummins (Eds.), *Disrupting finance* (pp. 33–50). Palgrave Pivot.

Azvine, B., Cui, Z., Majeed, B., & Spott, M. (2007). Operational risk management with real-time business intelligence. *BT Technology Journal*, *25*(1), 154–167. https://doi.org/10.1007/s10550-007-0017-5

Basel Committee on Banking Supervision. (2006). *International convergence of capital measurement and capital standards: A revised framework, comprehensive version*. Bank for International Settlements.

Basel Committee on Banking Supervision. (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Bank for International Settlements.

Basel Committee on Banking Supervision. (2020). *Governors and Heads of Supervision announce deferral of Basel III implementation to increase operational capacity of banks and supervisors to respond to Covid-19*. Bank for International Settlements.

Bouveret, A. (2019). Estimation of losses due to cyber risk for financial institutions. *Journal of Operational Risk*, *14*(2), 1–20. https://doi.org/10.21314/JOP.2019.224

Bromiley, P., McShane, M., Nair, A., & Rustambekov, E. (2015). Enterprise risk management: review, critique, and research directions. *Long Range Planning*, *48*(4), 265–276. https://doi.org/10.1016/j.lrp.2014.07.005

Buehler, K., Freeman, A., Hulme, R. (2008). *The risk revolution* (Working Paper).

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234.

Cao, L., Dong, X., & Zheng, Z. (2016). E-NSP: Efficient negative sequential pattern mining. *Artificial Intelligence*, *235*, 156–182. https://doi.org/10.1016/j.artint.2016.03.001

Caron, F., Vanthienen, J., & Baesens, B. (2012). Rule-based business process mining: Applications for management [Paper presentation]. 1st International Symposium on Management Intelligent Systems, Salamanca., IS-MiS 2012.

Caron, F., Vanthienen, J., & Baesens, B. (2013). A comprehensive investigation of the applicability of process mining techniques for enterprise risk management. *Computers in Industry*, *64*(4), 464–475. https://doi.org/10.1016/j.compind.2013.02.001

Centers for Disease Control and Prevention. (2017). *Historical mine disasters*. https://www.cdc.gov/niosh/mining/statistics/minedisasters.html

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, *83*(3), 735–776. https://doi.org/10.1111/jori.12059

Choi, T.-M., Chan, H. K., & Yue, X. (2017). Recent development in big data analytics for business operations and risk management. *IEEE Transactions on Cybernetics*, *47*(1), 81–92.

Choi, T.-M., & Lambert, J. H. (2017). Advances in risk analysis with big data. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *37*(8), 1435–1442.

Chollet, F. (2018). *Deep learning with Python*. Manning Publications Co.

Chu, C. Y., Park, K., & Kremer, G. E. (2020). A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. *Advanced Engineering Informatics*, *45*, 101053. https://doi.org/10.1016/j.aei.2020.101053

COSO. (2004). *Enterprise risk management – integrated framework*. Committee of Sponsoring Organizations of the Treadway Commission.

COSO. (2017). *Enterprise risk management – integrating with strategy and performance*. Committee of Sponsoring Organizations of the Treadway Commission.

COSO. (2020). *Guidance on enterprise risk management*. Committee of Sponsoring Organizations of the Treadway Commission. Retrieved 23 March from https://www.coso.org/Pages/erm.aspx

da Silva, J. B. N., Senna, P., Chousa, A., & Coelho, O. (2020). Data mining and operations research techniques in supply chain risk management: A bibliometric study. *Brazilian Journal of Operations & Production Management*, *17*(3), 14. https://doi.org/10.14488/BJOPM.2020.029

Dickinson, G. (2001). Enterprise risk management: Its origins and conceptual foundation. *The Geneva Papers on Risk and Insurance – Issues and Practice*, *26*(3), 360–366. https://doi.org/10.1111/1468-0440.00121

Dicuonzo, G., Galeone, G., Zappimbulso, E., & Dell'Atti, V. (2019). Risk management 4.0: The role of big data analytics in the bank sector. *International Journal of Economics and Financial Issues*, *9*(6), 40–47. https://doi.org/10.32479/ijefi.8556

Dong, M., Nassif, A. B., & Li, B. (2019). A data-driven residential transformer overloading risk assessment method. *IEEE Transactions on Power Delivery*, *34*(1), 387–396. https://doi.org/10.1109/TPWRD.2018.2882215

Dutta, K., & Perry, J. (2006). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital.

Eckle, P., & Burgherr, P. (2013). Bayesian data analysis of severe fatal accident risk in the oil chain. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *33*(1), 146–160.

El Bouchti, A., Chakroun, A., Abbar, H., & Okar, C. (2017). Fraud detection in banking using deep reinforcement learning [Paper presentation]. 7th International Conference on Innovative Computing Technology, INTECH 2017.

Embrechts, P., Mizgier, K. J., & Chen, X. (2018). Modeling operational risk depending on covariates: An empirical investigation. *Journal of Operational Risk*, *13*(3), 17–46. https://doi.org/10.21314/JOP.2018.212

Eryilmaz, M. (2018). Enterprise risk management: Past, today and future. In M. Rusev, E. Straus, C. Avcikurt, A. Soykan, & B. Parlak (Eds.), *Social science researches in the globalizing world* (pp. 244–254). St. Kliment Ohridski University Press.

European Commission Non-Financial Reporting Directive. (2014). Directive 2014/95/EU C.F.R. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095

Evans, J. (2020). *Pandemic briefing: Risk management implications of coronavirus (COVID-19) – for management*. https://actuaries.logicaldoc.cloud/download-ticket?ticketId=82acb23f-9164-4e4c-8621-72f4161d1a13

Fagundes, M. V. C., Teles, E. O., Vieira de Melo, S. A. B., & Freires, F. G. M. (2020). Decision-making models and support systems for supply chain risk: Literature mapping and future research agenda. *European Research on Management and Business Economics*, *26*(2), 63–70. https://doi.org/10.1016/j.iedeen.2020.02.001

Fang, L., Wei, L., Yi, G., & Jiang, L. (2017). Research of potential safety hazard investigation and risk control system for mine enterprise [Paper presentation]. 2nd IEEE International Conference on Computational Intelligence and Applications, ICCIA 2017.

Fedushko, S., Ustyianovych, T., & Gregus, M. (2020). Real-time high-load infrastructure transaction status output prediction using operational intelligence and big data technologies. *Electronics*, *9*(4), 668. https://doi.org/10.3390/electronics9040668

Forrester, J. W. (1961). *Industrial dynamics*. M.I.T. Press.

Groth, K. M., Denman, M. R., Darling, M. C., Jones, T. B., & Luger, G. F. (2020). Building and using dynamic risk-informed diagnosis procedures for complex system accidents. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *234*(1), 193–207. https://doi.org/10.1177/1748006X18803836

Groth, K. M., & Bensi, M. (2018). Commentary on use of model-augmented data analytics for improved operational efficiency of nuclear power plants [Paper presentation]. 14th Probabilistic Safety Assessment and Management, PSAM 2018.

Gudivada, V. N. (2017). Data analytics: Fundamentals. In A. Apon, M. Chowdhury, & K. Dey (Eds.), *Data analytics for intelligent transportation systems* (pp. 31–67). Elsevier Inc.

Guo, S., Liu, Y., He, J., Yang, T., He, G., Guo, S., Zhao, J., & Yang, H. (2017). A spot check information management system of wind farm [Paper presentation]. ASME 2017 Power Conference Joint with ICOPE 2017, POWER 2017-ICOPE 2017, Collocated with the ASME 2017 11th International Conference on Energy Sustainability, the ASME 2017 15th International Conference on Fuel Cell Science, Engineering and Technology, and the ASME 2017 Nuclear Forum. https://doi.org/10.1115/POWER-ICOPE2017-3308

Güven, S., Murthy, K., Shwartz, L., & Paradkar, A. (2016). Towards establishing causality between change and incident [Paper presentation]. 2016 IEEE/IFIP Network Operations and Management Symposium, NOMS 2016.

Hajakbari, M. S., & Minaei-Bidgoli, B. (2014). A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. *Journal of Loss Prevention in the Process Industries*, *32*, 443–453. https://doi.org/10.1016/j.jlp.2014.10.013

Hamedifar, H., Spitzenberger, C., Stahl, C., Brown, A., Nilberg, B., Demay, V., & Aspholm, O. (2015). Terminal and transportation risk assessment for LNG export in North America [Paper presentation]. Offshore Technology Conference, 2015, OTC 2015. https://doi.org/10.4043/26020-MS

Hao, C. (2013). ARMS: An advanced operational risk management system for commercial banks [Paper presentation]. 2013 3rd International Conference on Intelligent System Design and Engineering Applications, Hong Kong. ISDEA 2013

Harmantzis, F., & Malek, M. (2004). Security risk analysis and evaluation [Paper presentation]. 2004 IEEE International Conference on Communications, Paris.

Hassini, N., & Zouairi, S. (2011). Bayesian decision-making for industrial production facilities and processing [Paper presentation]. Saudi International Electronics, Communications and Photonics Conference 2011, Riyadh. SIECPC 2011

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Haustein, K., Widzyk-Capehart, E., Beitz, A., & Hainsworth, D. (2008). *Nexsys^{TM}: A Realtime Risk Management System for Mine Safety* [Paper presentation]. Queensland Mining Industry Health & Safety Conference 2008.

Hayne, K. M. (2019). *Final report: Royal commission into misconduct in the banking, superannuation and financial services industry*.

He, B. H., & Song, G. F. (2009). Knowledge management and data mining for supply chain risk management [Paper presentation]. International Conference on Management and Service Science, Wuhan., MASS 2009.

Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, *122*, 104492. https://doi.org/10.1016/j.ssci.2019.09.015

Hosseini, S., & Ivanov, D. (2020). Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review. *Expert Systems with Applications*, *161*, 113649. https://doi.org/10.1016/j.eswa.2020.113649

Huang, Z. M., Cong, Q. S., & Hu, J. B. (2012). Information system risk auditing model based on process mining [Paper presentation]. 2012 19th Annual International Conference on Management Science and Engineering, Dallas, TX., ICMSE 2012.

IBISWorld. (2021). Global *biggest industries by revenue* in 2021. IBISWorld https://www.ibisworld.com/global/industry-trends/biggest-industries-by-revenue/

International Actuarial Association. (2021). About the IAA. *International actuarial association*. Retrieved 14 July from https://www.actuaries.org/iaa/IAA/About_the_IAA/IAA/About_the_IAA/About_the_IAA/Vision_Mission.aspx?hkey=5f039c31-0926-46b1-becc-14dff02287b7

ISO (2009). *Risk management - Principles and guidelines*. (Vol. ISO 31000:2009, pp. 24). International Organization for Standardization.

ISO (2016). *Petroleum, petrochemical and natural gas industries - Collection and exchange of reliability and maintenance data for equipment*. (Vol. ISO 14224:2016, pp. 272). International Organization for Standardization.

ISO (2018). *Risk management - Guidelines*. (Vol. ISO 31000:2018, pp. 16). International Organization for Standardization.

Jans, M., Lybaert, N., & Vanhoof, K. (2010a). Data mining and economic crime risk management. *Surveillance technologies and early warning systems: Data mining applications for risk detection* (pp. 205–227). IGI Global.

Jans, M., Lybaert, N., & Vanhoof, K. (2010b). Internal fraud risk reduction: Results of a data mining case study. *International Journal of Accounting Information Systems*, *11*(1), 17–41. https://doi.org/10.1016/j.accinf.2009.12.004

Johnson, T. (2010). *Conceptual mapping of risk management to data mining* [Paper presentation]. 3rd International Conference on Emerging Trends in Engineering and Technology, Goa., ICETET 2010.

Johnston, J., & Guichard, A. (2015). New findings in drilling and wells using big data analytics [Paper presentation]. Offshore Technology Conference 2015, OTC 2015. https://doi.org/10.4043/26021-MS

Kaplan Higher Education. (2019). Operational *risk m*anagement (FIN219) [Course Content].

Kenett, R. S., & Salini, S. (2010). Measures of association applied to operational risks. In R. S. Kenett & Y. Raanan (Eds.), *Operational risk management: A practical approach to intelligent data analysis* (pp. 149–167). John Wiley and Sons.

KPMG US & The Risk Management Association. (2018). *Operational risk management excellence report (Executive report)*.

KPMG. (2021). *Risks and opportunities for mining*.

Lacković, I. D., Kovšca, V., & Vincek, Z. L. (2020). A review of selected aspects of big data usage in banks' risk management. *Journal of Information and Organizational Sciences*, *44*(2), 317–330. https://doi.org/10.31341/jios.44.2.7

Lambrigger, D., Shevchenko, P., & Wüthrich, M. (2007). The quantification of operational risk using internal data, relevant external data and expert opinions. *The Journal of Operational Risk*, *2*(3), 3–27. https://doi.org/10.21314/JOP.2007.030

Lee, D., Kim, J. G., & Kim, K. J. (2013). A study on abnormal event correlation analysis for convergence security monitor. *Cluster Computing*, *16*(2), 219–227. https://doi.org/10.1007/s10586-011-0191-6

Leidner, J. L., & Schilder, F. (2010). Hunting for the black swan: Risk mining from text [Paper presentation]. 48th Annual Meeting of the Association for Computational Linguistics, Uppsala., ACL 2010.

Leite, S. R. (2019). Prediction analysis of forced shutdowns in Brazilian Type I hydropower plants [Paper presentation]. 2019 14th Iberian Conference on Information Systems and Technologies, New York.

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29. https://doi.org/10.3390/risks7010029

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Li, H., Lan, N. W., & Huang, X. N. (2020). *Application research of fault diagnosis in conventional island of nuclear power plant based on support vector machine* [Paper presentation]. 4th International Symposium on Software Reliability, Industrial Safety, Cyber Security and Physical Protection of Nuclear Power Plant, ISNPP 2019.

Liang, W., Sari, A., Zhao, G., McKinnon, S. D., & Wu, H. (2020). Short-term rockburst risk prediction using ensemble learning methods. *Natural Hazards*, *104*(2), 1923–1946. https://doi.org/10.1007/s11069-020-04255-7

Lien, C. C. (2012). The application of crisp and fuzzy decision trees to monitor insurance customer database. *Information*, *15*(9), 3871–3876.

Lien, C. C., Ho, C. C., & Tsai, Y. M. (2011). Applying fuzzy decision tree to infer abnormal accessing of insurance customer data [Paper presentation]. 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011, Jointly with the 2011 7th International Conference on Natural Computation, Shanghai., ICNC'11.

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, *45*(2), 175–194. https://doi.org/10.1177/0312896219877678

Liu, T. B., & Liu, S. P. (2016). Fraud detection model & application for credit card acquiring business based on data mining technology [Paper presentation]. Proceedings of the 2016 4th International Conference on Electrical & Electronics Engineering and Computer Science, Paris.

Marquez, A. C., Del Castillo, A. C., & Fernandez, J. F. G. (2020). Integrating artificial intelligent techniques and continuous time simulation modelling. Practical predictive analytics for energy efficiency and failure detection. *Computers in Industry*, *115*, 15.

Mazumder, R. K., Salman, A. M., & Li, Y. (2021). Failure risk analysis of pipelines using data-driven machine learning algorithms. *Structural Safety*, *89*, 102047. https://doi.org/10.1016/j.strusafe.2020.102047

McKibbin, W., & Fernando, R. (2021). The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Economic Papers*, *20*(2), 1–30. https://doi.org/10.1162/asep_a_00796

McKinsey & Company, & Operational Riskdata eXchange Association. (2017). *The future of operational risk*.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools* (Revised ed.). Princeton University Press.

Middleton, S. E., & Sabeur, Z. A. (2011). *Knowledge-based service architecture for multi-risk environmental decision support applications*. Environmental Software Systems: Frameworks of Eenvironment.

Milana, D., Darena, M. S., Bettio, N., Cerruti, C., Siliprandi, G., Fidanzi, A., Cerioli, P., Silvestri, G., Tarasconi, F., Caserio, M., Botros, M., & Gabrielli, M. L. (2019). Natural language understanding for safety and risk management in oil and gas plants [Paper presentation]. Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019. https://doi.org/10.2118/197668-MS

Milkau, U., & Bott, J. (2018). Active management of operational risk in the regimes of the "Unknown": What can machine learning or heuristics deliver? *Risks, 6*(2), 41.

Mišić, V. V., & Perakis, G. (2020). Data analytics in operations management: A review. *Manufacturing & Service Operations Management*, *22*(1), 158–169. https://doi.org/10.1287/msom.2019.0805

Mittnik, S., & Starobinskaya, I. (2010). Modeling Dependencies in Operational Risk with Hybrid Bayesian Networks. *Methodology and Computing in Applied Probability*, *12*(3), 379–390. https://doi.org/10.1007/s11009-007-9066-y

Modarres, M., Kaminskiy, M. P., & Krivtsov, V. (2016). *Reliability engineering and risk analysis: A practical guide* (3rd ed.). Taylor & Francis Group.

Mohaghegh, Z., Kazemi, R., & Mosleh, A. (2009). Incorporating organizational factors into probabilistic risk assessment (PRA) of complex socio-technical systems: A hybrid technique formalization. *Reliability Engineering & System Safety*, *94*(5), 1000–1018. https://doi.org/10.1016/j.ress.2008.11.006

Monish, H., & Pandey, A. C. (2020). A comparative assessment of data mining algorithms to predict fraudulent firms [Paper presentation]. 10th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2020.

Moura, R., Beer, M., Patelli, E., Lewis, J., & Knoll, F. (2017). Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies. *Safety Science*, *99*, 196–214. https://doi.org/10.1016/j.ssci.2017.05.001

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian networks in R with applications in systems biology* (1st ed.). Springer.

Nagashree, N., Tejasvi, R., & Swathi, K. C. (2018). An early risk detection and management system for the cloud with log parser. *Computers in Industry*, *97*, 24–33.

Nateghi, R., & Aven, T. (2021). Risk analysis in the age of big data: The promises and pitfalls. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *41*(10), 1751–1758.

Neil, M., Fenton, N., & Tailor, M. (2005). Using Bayesian networks to model expected and unexpected operational losses. *Risk Analysis*, *25*(4), 963–972. https://doi.org/10.1111/j.1539-6924.2005.00641.x

Neil, M., Häger, D., & Andersen, L. B. (2009). Modeling operational risk in financial institutions using hybrid dynamic Bayesian networks. *The Journal of Operational Risk*, *4*(1), 3–33. https://doi.org/10.21314/JOP.2009.057

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

Nugent, T., & Leidner, J. L. (2017). Risk mining: Company-risk identification from unstructured sources [Paper presentation]. 16th IEEE International Conference on Data Mining Workshops, ICDMW 2016.

Nwafor, C. N., Nwafor, O. Z., & Onalo, C. (2019). The use of business intelligence and predictive analytics in detecting and managing occupational fraud in Nigerian banks. *Journal of Operational Risk*, *14*(3), 95–120.

Ocelewicz, L., Lewis, J., Steele, C. (2021). Conduct risk: Delivering an effective framework. https://home.kpmg/uk/en/home/insights/2017/09/conduct-risk-delivering-an-effective-framework.html

Onoda, T., Ito, N., & Yamasaki, H. (2006). Unusual condition mining for risk management of hydroelectric power plants [Paper presentation]. 6th IEEE International Conference on Data Mining, Hong Kong.

Onoda, T., Ito, N., & Yamasaki, H. (2009). Interactive trouble condition sign discovery for hydroelectric power plants [Paper presentation]. 15th International Conference on Neuro-Information Processing, Auckland., ICONIP 2008.

Operational Riskdata eXchange Association. (2020). *Annual banking loss report*. Operational Riskdata eXchange Association.

O'Shea, N., Pence, J., Mohaghegh, Z., & Kee, E. (2015). Physics of failure, predictive modeling & data analytics for LOCA frequency [Paper presentation]. 61st Annual Reliability and Maintainability Symposium, RAMS 2015.

Palshikar, G. K., & Apte, M. (2013). Financial security against money laundering: A survey. In B. Akhgar & H. Arabnia (Eds.), *Emerging trends in ICT security* (pp. 577–590). Elsevier Inc.

Parhizi, S., Shahrabi, J., & Pariazar, M. (2009). A new accident investigation approach based on data mining techniques. *Journal of Applied Sciences*, *9*(4), 731–737. https://doi.org/10.3923/jas.2009.731.737

Pence, J., Farshadmanesh, P., Kim, J., Blake, C., & Mohaghegh, Z. (2020). Data-theoretic approach for socio-technical risk analysis: Text mining licensee event reports of U.S. nuclear power plants. *Safety Science*, *124*, 104574. https://doi.org/10.1016/j.ssci.2019.104574

Pence, J., Mohaghegh, Z., Ostroff, C., Kee, E., Yilmaz, F., Grantom, R., & Johnson, D. (2014). Toward monitoring organizational safety indicators by integrating probabilistic risk assessment, socio-technical systems theory, and big data analytics [Paper presentation]. 12th International Probabilistic Safety Assessment and Management Conference, PSAM 2014.

Pence, J., Sakurahara, T., Zhu, X. F., Mohaghegh, Z., Ertem, M., Ostroff, C., & Kee, E. (2019). Data-theoretic methodology and computational platform to quantify organizational factors in socio-technical risk analysis. *Reliability Engineering & System Safety*, *185*, 240–260. https://doi.org/10.1016/j.ress.2018.12.020

Persona, A., Battini, D., Faccio, M., Bevilacqua, M., & Ciarapica, F. E. (2006). Classification of occupational injury cases using the regression tree approach. *International Journal of Reliability, Quality and Safety Engineering*, *13*(02), 171–191. https://doi.org/10.1142/S0218539306002197

Peters, G. W., Clark, G., Thirlwell, J., & Kulwal, M. (2018). Global perspectives on operational risk management and practice: A survey by the Institute of Operational Risk (IOR) and the Center for Financial

Professionals (CeFPro). *Journal of Operational Risk*, *13*(4), 47–88. https://doi.org/10.21314/JOP.2018.215

Peters, G. W., Shevchenko, P. V., & Wüthrich, M. V. (2009). Dynamic operational risk: Modeling dependence and combining different sources of information. *The Journal of Operational Risk*, *4*(2), 69–104. https://doi.org/10.21314/JOP.2009.059

Pika, A., van der Aalst, W. M. P., Fidge, C. J., ter Hofstede, A. H. M., & Wynn, M. T. (2013). Profiling event logs to configure risk indicators for process delays [Paper presentation]. 25th International Conference on Advanced Information Systems Engineering, Valencia., CAiSE 2013.

PwC. (2017). Risk in review: Managing risk from the front line.

Qazi, A., Quigley, J., & Dickson, A. (2015). Supply *chain risk management*: Systematic literature review and a conceptual framework for capturing interdependencies between risks [Paper presentation]. 5th International Conference on Industrial Engineering and Operations Management, IEOM 2015.

Rahmes, M., Wilder, K., Yates, H., & Fox, K. (2013). Near real time discovery and conversion of open source information to a reward matrix [Paper presentation]. 17th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, Florida., WMSCI 2013.

Rajesh, R. (2020a). A grey-layered ANP based decision support model for analyzing strategies of resilience in electronic supply chains. *Engineering Applications of Artificial Intelligence*, *87*, 103338. https://doi.org/10.1016/j.engappai.2019.103338

Rajesh, R. (2020b). A novel advanced grey incidence analysis for investigating the level of resilience in supply chains. *Annals of Operations Research*, *308*(1-2), 441–490. https://doi.org/10.1007/s10479-020-03641-5

Rasmussen, J. (1990). Human error and the problem of causality in analysis of accidents. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *327*(1241), 449–462.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2–3), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Reason, J. (1990). *Human error*. Cambridge University Press.

Reason, J. (2000). Human error: Models and management. *BMJ (Clinical Research ed.)*, *320*(7237), 768–770.

Risk Leadership Network. (2021). Why join the *risk leadership network*? Risk Leadership Network. Retrieved 30 June from https://www.riskleadershipnetwork.com/why-join

Rodriguez-Ulloa, R. (2018). Systemic methodology for risks evaluation and management in the energy and mining sectors (SYSMEREM-EMS) using Bayesian networks. *Journal of Decision Systems*, *27*(sup1), 191–200. https://doi.org/10.1080/12460125.2018.1468157

Saaty, T. L., & Peniwati, K. (2008). *Group decision making: Drawing out and reconciling differences*. RWS Publications.

Safe Work Australia. (2020). *Work-related traumatic injury fatalities, Australia*. S. W. Australia.

Sanchez-Pi, N., Martí, L., Molina, J. M., & Garcia, A. C. B. (2014). High-level information fusion for risk and accidents prevention in pervasive oil industry environments. *Highlights of practical applications of heterogeneous multi-agent systems: The Paams Collection*.

Sanchez-Pi, N., Martí, L., Molina, J. M., & Garcia, A. C. B. (2015). Contextual pattern discovery in ambient intelligent application. *International Journal of Imaging and Robotics*, *15*(4), 165–178.

Sanford, A., & Moosa, I. (2012). A Bayesian network structure for operational risk modelling in structured finance operations. *Journal of the Operational Research Society*, *63*(4), 431–444. https://doi.org/10.1057/jors.2011.7

Sanford, A., & Moosa, I. (2015). Operational risk modelling and organizational learning in structured finance operations: A Bayesian network approach. *Journal of the Operational Research Society*, *66*(1), 86–115. https://doi.org/10.1057/jors.2013.49

Satoh, N., & Samejima, M. (2019). Risk words suggestion for information security audit by Bayesian inference. *Electronics and Communications in Japan*, *102*(1), 42–48. https://doi.org/10.1002/ecj.12133

SCImago (n.d). SJR – SCImago Journal & Country Rank. Retrieved 15 December 2019 from http://www.scimagojr.com

Setiono, R., Mues, C., & Baesens, B. (2006). Risk management and regulatory compliance: A data mining framework based on neural network rule extraction [Paper presentation]. 27th International Conference on Information Systems, Milwaukee, WI., ICIS 2006.

Shafiee, M., Animah, I., Alkali, B., & Baglee, D. (2019). Decision support methods and applications in the upstream oil and gas sector. *Journal of Petroleum Science and Engineering*, *173*, 1173–1186. https://doi.org/10.1016/j.petrol.2018.10.050

Shayboun, M., Kifokeris, D., & Koch, C. (2020). Machine learning for analysis of occupational accidents registration data [Paper presentation]. 36th Annual Conference on Association of Researchers in Construction Management, ARCOM 2020.

Shevchenko, P. V. (2015). Operational risk. In H. K. Baker & G. Filbeck (Eds.), *Investment risk management* (pp. 119–140). Oxford University Press.

Silva, J. F., & Jacinto, C. (2012). Finding occupational accident patterns in the extractive industry using a systematic data mining approach. *Reliability Engineering & System Safety*, *108*, 108–122. https://doi.org/10.1016/j.ress.2012.07.001

Sjöblom, O., Murtola, L. M., Heimonen, J., Kauhanen, L., Laippala, V., Lundgren-Laine, H., Salakoski, T., & Salanterä, S. (2013). Using cluster analysis to identify weak signals of lethal trends in aviation and healthcare documentation. *International Journal of Networking and Virtual Organisations*, *13*(1), 66–80. https://doi.org/10.1504/IJNVO.2013.058442

Standards Australia. (1995). *AS/NZS 4360-1995 risk management (Vol. AS/NZS 4360-1995)*.

Stojadinovic, S., Svrkota, I., Petrovic, D., Denic, M., Pantovic, R., & Milic, V. (2012). Mining injuries in Serbian underground coal mines – A 10-year study. *Injury*, *43*(12), 2001–2005. https://doi.org/10.1016/j.injury.2011.08.018

The Institute of Internal Auditors. (2013). *The three lines of defense in effective risk management and control*. https://na.theiia.org/standards-guidance/Public%20Documents/PP%20The%20Three%20Lines%20of%20Defense%20in%20Effective%20Risk%20Management%20and%20Control.pdf

Thomson Reuters. (2010). *TIMELINE-Major offshore accidents in the global oil industry*. https://www.reuters.com/article/idUSN13272386

Tubis, A. A., Werbińska-Wojciechowska, S., Góralczyk, M., Wróblewski, A., & Ziętek, B. (2020). Cyber-attacks risk analysis method for different levels of automation of

mining processes in mines based on fuzzy theory use. *Sensors*, *20*(24), 7210–7223. https://doi.org/10.3390/s20247210

Udo, M., & Jürgen, B. (2018). Active management of operational risk in the regimes of the "unknown": what can machine learning or heuristics deliver? *Risks*, *6*(2).

Urabe, Y., Yamanishi, K., Tomioka, R., & Iwai, H. (2011). Real-*time change-point detection using sequentially discounting normalized maximum likelihood coding* [Paper presentation]. Advances in Knowledge Discovery and Data Mining, Pt Ii: 15th Pacific-Asia Conference, Berlin., Pakdd 2011.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Vishnu, C. R., Sridharan, R., & Kumar, P. N. R. (2019). Supply chain risk management: Models and methods. *International Journal of Management and Decision Making*, *18*(1), 31–75. https://doi.org/10.1504/IJMDM.2019.096689

Wang, Y., Li, G., Li, J., & Zhu, X. (2018). Comprehensive identification of operational risk factors based on textual risk disclosures [Paper presentation]. 6th International Conference on Information Technology and Quantitative Management, ITQM 2018.

Wang, Y., Li, J., & Zhu, X. (2017). A method of estimating operational risk: loss distribution approach with piecewise-defined frequency dependence [Paper presentation]. 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Wei, L., Li, G., Zhu, X., Sun, X., & Li, J. (2019). Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures. *Energy Economics*, *80*, 452–460. https://doi.org/10.1016/j.eneco.2019.01.020

Wei, Y. C., Wu, W. C., & Chu, Y. C. (2018). Performance evaluation of the recommendation mechanism of information security risk identification. *Neurocomputing*, *279*, 48–53. https://doi.org/10.1016/j.neucom.2017.05.106

World Economic Forum. (2021). Fourth industrial revolution. Retrieved 25 June from https://www.weforum.org/focus/fourth-industrial-revolution

Wu, Q., Liu, Y. Z., Liu, D. H., & Zhou, W. F. (2011). prediction of floor water inrush: The application of GIS-based AHP vulnerable index method to Donghuantuo Coal Mine, China. *Rock Mechanics and Rock Engineering*, *44*(5), 591–600. https://doi.org/10.1007/s00603-011-0146-5

Wu, W. (2020). Construction of ecological monitoring and early warning system in coal mine based on big data analysis. *Fresenius Environmental Bulletin*, *29*(5), 3564–3570.

Wu, Y., Wang, X., & Zhang, Z. (2015). A high risk association rule mining algorithm for the petroleum industry. *Journal of Information and Computational Science*, *12*(12), 4633–4642. https://doi.org/10.12733/jics20106385

Wu, Z. G., Zhang, X. Y., Xiao, C. G., & Chen, W. (2019). Applications of data mining in conventional island of nuclear power plant [Paper presentation]. 3rd International Symposium on Software Reliability, Industrial Safety, Cyber Security and Physical Protection of Nuclear Power Plant, ISNPP 2018.

Xiaorong, C., Yan, W., & Xin, G. (2009). Network security risk assessment based on association rules [Paper presentation]. 2009 4th International Conference on Computer Science and Education, Nanning., ICCSE 2009.

Xiaoyun, W., & Danyue, L. (2010). Hybrid outlier mining algorithm based evaluation of client moral risk in insurance company [Paper presentation]. 2010 2nd IEEE International Conference on Information Management and Engineering, Chengdu., ICIME 2010. https://doi.org/10.1109/ICIME.2010.5478070

Xie, X. C., Fu, G., Xue, Y. J. Y., Zhao, Z. Q., Chen, P., Lu, B. J., & Jiang, S. (2019). Risk prediction and factors risk analysis based on IFOA-GRNN and apriori algorithms: Application of artificial intelligence in accident prevention. *Process Safety and Environmental Protection*, *122*, 169–184. https://doi.org/10.1016/j.psep.2018.11.019

Yin, Q., Yang, J., Tyagi, M., Zhou, X., Hou, X., & Cao, B. (2021). Field data analysis and risk assessment of gas kick during industrial deepwater drilling process based on supervised learning algorithm. *Process Safety and Environmental Protection*, *146*, 312–328. https://doi.org/10.1016/j.psep.2020.08.012

Zhang, K., Wu, X., Niu, R., Yang, K., & Zhao, L. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. *Environmental Earth Sciences*, *76*(11), 405. https://doi.org/10.1007/s12665-017-6731-5

Zhu, X., Li, J., Chen, J., Yanghuo, Y., Gao, L., Feng, J., Wu, D., & Xie, Y. (2014). A nonparametric operational risk modeling approach based on Cornish-Fisher expansion. *Discrete Dynamics in Nature and Society*, *2014*, 1–8. https://doi.org/10.1155/2014/839731

Appendices

Appendix A: Systematic literature review methodology details

Details of the SLR Searches

Search ID	Search strategy	Industry	Database	Search date	References retrieved
SEA111[ab]	TITLE-ABS-KEY (("data analy*" OR "machine learning" OR "big data" OR "artificial intelligence" OR "business intelligence" OR "data min*") AND ("risk manag*" OR "risk analy*" OR "risk framework" OR "operational risk") AND ("financial services" OR bank* OR "financ* industry" OR "financ* sector"))	FS	Scopus	09/02/2021	431
SEA121[a]	TITLE-ABS-KEY (("data analy*" OR "machine learning" OR "big data" OR "artificial intelligence" OR "business intelligence" OR "data min*") AND ("risk manag*" OR "risk analy*" OR "risk framework" OR "risk decision" OR "operational risk") AND ("energy" OR "mining"))	ENR	Scopus	09/02/2021	1,476
SEA211	TOPIC: (("data analy*" OR "machine learning" OR "big data" OR "artificial intelligence" OR "business intelligence" OR "data min*") AND ("risk manag*" OR "risk analy*" OR "risk framework" OR "risk decision" OR "operational risk") AND ("financial services" OR bank* OR "financ* industry" OR "financ* sector" OR "energy" OR "mining"))	FS, ENR	Web of Science	09/02/2021	631
SEA000	Various	General, FS, ENR	Google Scholar and Other	Various	35

[a]Because of the maximum character restriction when searching Scopus, the search strategy was split by industry, such that two searches (SEA111 and SEA121) were conducted in Scopus.
[b]The risk management search term "risk decision" was excluded from SEA111 on account of the maximum character restriction in Scopus. However, the effect of its exclusion on the number of references retrieved from the search was tested for and none was found.

Stage 1 Cleaning Criteria

Inclusion criteria	Exclusion criteria
(a) Study Objective: Application of data analytics technique(s) to the process of managing risks.	(a) Study Objective: • Broad discussion of a risk(s) with no application of a data analytics solution; • Effect of risks or risk management practices on an organisation's operations, stability, or financial performance; • Broad overview of the possible applications of data analytics, with risk management mentioned only briefly; • Specific modelling theory without practical application to a risk management situation; • Conference proceedings; or • Review of a journal.
(b) At least one risk analysed was one of the following: • Credit Risk (including bankruptcy, individual consumer credit scores); • Liquidity Risk; • Market Risk; • Operational Risk (including specific risk events, technology, system or cyber security risks); • Fraud; • Reputational Risk; • Safety Risk; • External Risk; • Emergency Risk or Crisis Management; • Strategic Risk; or • Supply Chain Risk.	(b) Highly specific and technical risk event, not outlined in *Inclusion Criterion (b)*.
(c) Industry(ies) of application was at least one of the following: • Financial Services (including banks, digital finance firms, or insurers); or • Energy (including mining, oil, gas, coal).	(c) Accessibility of Reference: • Full-text not available; or • Not written in English language.

Stage 2 Cleaning Criteria

Inclusion criteria	Exclusion criteria
(a) Application of data analytic(s) technique to any operational risk event/s (micro/specific or aggregate) at an organisational level.	(a) Non-operational risks, including: • Credit Risk (including bankruptcy, individual consumer credit scores); • Liquidity Risk; and • Market Risk. (b) Analysis of community-wide, external risks, including: • Community emergency management for natural disasters and terrorist attacks; and • Traffic management.