The International Journal of Digital Accounting Research
Vol. 23, 2023, pp. 69-96
ISSN: 2340-5058

The value relevance of digitalization disclosure in integrated reports: A South African perspective

Aneetha Sukhari. University of Johannesburg, South Africa, sukhaar@unisa.ac.za

Daniël Coetsee. University of Johannesburg, South Africa, dcoetsee@uj.ac.za

Abejide Ade-Ibijola. University of Johannesburg, South Africa, abejide@jbs.ac.za

Abstract. The inevitable disruptions in the Fourth Industrial Revolution necessitates that companies provide investors with digitalization disclosure in integrated reports. This paper investigated whether digitalization disclosure in integrated reports affects the share prices of South African listed companies. The relationship between digitalization disclosure and share prices is examined using the Ohlson (1995) Model through the application of panel data. A new proxy for the "other information" variable in the Ohlson (1995) Model was created for digitalization disclosure by developing a disclosure index to measure the scope of digitalization disclosure in integrated reports. The disclosure index was incorporated into a new text analysis software named the Fourth Industrial Revolution Disclosure Analysis Tool (4IRDAT), which uses algorithms based on natural language processing techniques to facilitate the content analysis of digitalization disclosure in integrated reports. Two scenarios were evaluated: including loss-making companies and excluding loss-making companies. The sample size, including loss-making companies and excluding loss-making companies, was 90 (270 observations) and 72 (216 observations), respectively, for three years from 2018 to 2020. It was established that there was an increase in digitalization disclosure over three years. The results indicated that digitalization disclosure had yet to be incorporated in the share price of South African listed companies for both scenarios. This study is indispensable to regulators, practitioners, standard setters, and academics because it provides empirical evidence on the value relevance of digitalization disclosure in integrated reports. This area has not been interrogated in a South African context.

Keywords: Integrated reports, digitalization disclosure, Ohlson (1995) model, natural language processing techniques, algorithms.

Submitted January 2023
Accepted July 2023

DOI: 10.4192/1577-8517-v23_4

1. INTRODUCTION

The World Economic Forum found that the Covid-19 pandemic has accelerated the Fourth Industrial Revolution (4IR) globally and will significantly influence how companies do business going forward (IoDSA, 2016; World Economic Forum, 2019). Technologies emerging in the 4IR will disrupt how companies generate value, resulting in the establishment of new or modified business models (Ibarra et al., 2018; Klingenberg et al., 2021; Li et al., 2017; Piccarozzi et al., 2018; Wagire et al., 2019). Companies are required to adapt rapidly to maintain their competitive advantage. Stakeholders expect companies to invest in digital technologies to survive and prosper (Hossnofsky & Junge, 2019). Implementing specific technologies in a company would be of interest to investors' decision-making on whether to buy, sell or hold shares in a particular company. It would also give investors an indication as to whether a company would survive in the industry in the long term, which could influence the decision to buy, sell or hold shares in a particular company. Corporate reports, such as integrated reports, would be a source of information for investors to evaluate digital activity in a company.

The increase in corporate disclosure in recent years has culminated in large amounts of information for capital market participants in the form of texts, numbers, images, and tables (Dyer et al., 2017; Lewis & Young, 2019). The information available in corporate reports is unstructured data. Unstructured data is characterized as data with no defined format and cannot be easily retrieved consistently using automated techniques across companies and over time (Köseoğlu, 2022; Lewis & Young, 2019). This unstructured data creates difficulty in analyzing corporate reports over time across companies and industries (Türegün, 2019). Lewis & Young (2019) highlight the need to process the information in corporate reports into structured data for deeper analysis for users of corporate reports. The research problem identified is that digitalization disclosure in integrated reports is unstructured data, which makes it challenging to evaluate across companies and periods, which decreases its usefulness for decision-making. Furthermore, digitalization disclosure is not required by an accounting framework, and there are no reporting guidelines for digitalization information, which impedes the measurement of digitalization disclosure in integrated reports. The research question answered in this paper is – Is the scope of digitalization disclosure in integrated reports value relevant for investors?

Prior studies assessed the value relevance of digitalization disclosure in annual reports, earnings forecasts, conference calls, corporate websites, integrated reports, investment announcements and statutory filings and found that digitalization disclosure is useful to investors (Chen & Srinivasan, 2019; Hossnofsky & Junge, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021). Furthermore, other studies discussed whether digitalization disclosure influences firm performance or not (Zeng et al., 2022). The effect of digitalization on the risk of share price crashes was also assessed (Jiang et al., 2022). The studies that were mentioned above, developed digitalization dictionaries, and made use of manual content analysis (Chen & Srinivasan, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021) or packaged software (Hossnofsky & Junge, 2019). Also, other researchers used deep learning models (Jiang et al., 2022) or text mining techniques (Zeng et al., 2022) to measure digitalization disclosure. The research gap identified that more research on digitalization disclosure and firm value was required. Therefore, this paper aims to build on prior research by developing a disclosure index and evaluating the value relevance of digitalization disclosure in integrated reports.

2. LITERATURE REVIEW

In this section, two aspects of prior literature are reviewed: firstly, the literature on the application of natural language processing techniques to assess non-financial information and secondly, the measurement of digitalization disclosure.

2.1. Measuring non-financial disclosure using natural language processing techniques

A new development in the assessment of non-financial reporting is the use of natural language processing techniques (Türegün, 2019). Natural language processing is a subset of artificial intelligence and uses computational techniques to extract meaning from structured or unstructured text (Chowdhary, 2020; Kao & Poteet, 2007). Natural language processing techniques are widely used in text mining (Kao & Poteet, 2007). Text mining is identifying and extracting knowledge from free or unstructured text (Kao & Poteet, 2007). Examples of text mining are information retrieval, text classification and clustering, entity, relation, and event extraction.

Prior studies have applied natural language processing techniques to assess the quality of non-financial reporting (El-Haj et al., 2019; Miura et al., 2021; Nakagawa et al., 2020). Nakagawa et al., (2020) used natural language processing techniques to evaluate three aspects of integrated reports of Japanese listed companies. Firstly, integrated reports which were considered to be of excellent quality were identified. Next, the descriptive words used in the excellent reports were extracted. The sample of integrated reports was analysed for the presence of these words. The results indicated that companies with excellent integrated reports focus on value-creation in the long term, whereas poor quality reports focus on business activities. Secondly, the topics required to be included in the integrated reports were identified. The sample of integrated reports was assessed for the presence of these topics. The results found that excellent integrated reports disclosed more information about customers, products, and sustainability, whereas poor quality reports disclosed incentives for management. Thirdly, words related to environmental, social and governance disclosure were identified. The authors found that excellent and poor-quality reports have improved environmental, social and governance disclosure. Miura et al., (2021) used text mining techniques to identify the presence of specific words to assess the quality of Japanese universities' integrated reports. Firstly, a word cloud was created to identify the most frequent words in the sample. Secondly, the words were clustered to group the topics. The authors concluded that "revenue" and "expenses" appeared most frequently in integrated reports of universities.

El-Haj et al. (2020) assessed the quantity, readability, and tone of non-financial information in a sample of annual reports of UK-listed companies. Annual reports have an inconsistent structure; for example, section headings, page numbers and the order of sections differ and are made available in PDF format, which is a barrier to automated analysis of large samples. The authors presented a procedure for retrieving and classifying non-financial information in annual reports by developing a software tool which uses algorithms. There are four steps in the process. The first step is to retrieve all the text from the PDF annual reports. Next, the table of contents of each annual report is isolated, and a list of sections expected to be in the annual reports is developed. The algorithm then detects the sections from each annual report's table of contents and identifies the page numbers of each section. The second step is to classify the sections into financial or non-financial information. The third step is to classify each aspect of non-financial information into subcomponents, such as performance highlights, statements from the board chair, management commentary, governance statements and remuneration reports. All remaining information is categorized under other residual information. In the fourth step, the text is processed, and the scores are generated. The quantity was calculated using word counts and page counts. The tone was measured using word counts of positive and negative words. The level of readability was calculated using readability indexes. The study concluded that non-financial information's quantity, tone and readability increased. The software tool El Haj et al. (2020) developed has several advantages. Firstly, it allows for large samples of PDF reports to be evaluated. Secondly, researchers can amend the keyword list if new reporting requirements are implemented. Thirdly, the software can be used for annual reports in different languages. There are two disadvantages of the software. Firstly, the information is double counted if a section starts halfway down the page. Secondly, the software cannot isolate tables, charts, pictures, or infographic information.

Prior studies explored the benefits and challenges of manual and computer-aided content analysis, such as natural language processing techniques (El-Haj et al., 2019; Lewis & Young, 2019; Matthies & Coners, 2015; Zhang et al., 2019). These studies have concluded that combining manual and computer-aided content analysis is more beneficial for analyzing unstructured data in corporate reports. Natural language processing can supplement further manual scrutiny once the data has been collated (El-Haj et al., 2019; Lewis & Young, 2019; Matthies & Coners, 2015). The main challenges of using computer-aided content analysis are the need for programming skills and the high costs of developing software (Zhang et al., 2019).

2.2. Measuring digitalization disclosure

Prior studies assessed the effect of the digitalization disclosure on firm value (Chen & Srinivasan, 2019; Hossnofsky & Junge, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021). Some studies used the digitalization disclosure in annual reports (Hossnofsky & Junge, 2019; Ricci et al., 2020), others used information in investment announcements, earnings forecasts, annual statutory filings and conference calls (Chen & Srinivasan, 2019; Lui & Ngai, 2019) and one study explored digitalization disclosure on online platforms (Salvi et al., 2021). Zeng et al. (2022) assessed if digitalization disclosure influences firm performance or another study assessed the risk of share price crashes (Jiang et al., 2022).

Chen & Srinivasan (2019) assessed the value relevance of digital activity in earnings forecasts, annual statutory filings and conference calls using manual content analysis. A list of words was developed based on six topics: artificial intelligence, analytics, big data, cloud computing, digitization, and machine learning. The number of mentions was counted, and a scoring system of 0 to 3 was applied, with 3 being a high number of mentions. The final score was used as a proxy in the regression. The sample was limited to companies outside the main business of technology from 2010 to 2017. The authors found increased digital activity and that companies with higher digital activity had a higher firm value.

Lui & Ngai (2019) investigated the effect of blockchain on firm value using an events study. The Fama & French (1996) model was applied to investment announcements from 2015 to 2018 for 77 companies. The investment announcements were searched manually for information about engagement on blockchain. Thereafter, the abnormal return was calculated using regression analysis. The authors found that adopting blockchain technology positively affected a company’s market value in the long term.

Hossnofsky and Junge (2019) assessed whether analysts reacted positively to digitalization disclosure in annual reports of German-listed companies from 2006 to 2017. A digitalization dictionary was developed from the results of a survey, and the most frequent words were identified and included in the digitalization dictionary. A software package called Linguistic Inquiry Word Count (LIWC) was used to analyze the text in annual reports and provide the word counts. The study concluded that analysts react positively to digitalization in the long term.

Ricci et al. (2020) applied the Ohlson (1995) Model to assess whether digitalization disclosure in annual reports impacted firm value. The sample consisted of 75 Italian-listed companies from 2011 to 2017. The digitalization dictionary developed by Hossnofsky and Junge (2019) was applied in the manual content analysis of the annual reports. The number of words relating to digitalization was counted from the annual reports, specifically from the management commentaries and letters to shareholders. The score was calculated based on the total number of words counted divided by the total number of words in the digitalization dictionary. A proxy for digitalization-related disclosure was obtained for each firm-year observation. The authors found a positive association between the stock market valuation and the level of digitalization-related disclosure.

Salvi et al. (2021) evaluated if digitalization disclosure in corporate websites affected the share prices of 114 internationally listed companies using Tobin’s Q. A disclosure index was developed by identifying aspects from prior research and classifying the items into macro-components. Twenty-three items were identified under the following categories: instruments of digital communication; e-commerce; data management; information about digitalization; and investments in digitalization. The content analysis of the websites was conducted manually. A score of 1 was awarded if the item was disclosed and 0 if not. The total score ranged from 0 to a maximum of 23. The study concluded that digitalization disclosure had a positive effect on firm value.

Zeng et al. (2022) explored if digitalization disclosure in annual reports affected firm performance. Text-mining techniques were adopted to measure digitalization by using the following steps. Firstly, the authors developed a list of keywords related to digitalization. Secondly, all the keywords were extracted from the annual reports of a sample of companies from 2012 to 2019. Thirdly, negative expressions and information unrelated to the digitalization of the business were removed. Next, the degree of digitalization was calculated by awarding a score of 1 if the keyword appears and 0 otherwise. Thereafter, the sum of all keywords' value was calculated as each company's final score. The article found a low level of digitalization in the sample of Chinese-listed companies, and digitalization positively affected firm performance.

Jiang et al. (2022) explored the effect of digitalization on the risk of share price crashes. The sample included all Chinese-listed companies ranging from 2007 to 2020. An indicator of digital transformation was developed by identifying the level of investment in digital intangible assets. A digital dictionary was constructed from policy documents, and a deep learning model was trained and used to extract the digital information from the annual reports in the sample. The final score was calculated as the proportion of digital intangible assets to total assets. The authors found that digitalization reduces the risk of share price crashes.

In summary, prior studies used self-developed disclosure indices with digitalization dictionaries and manual content analysis, packaged software, deep learning models and text-mining techniques to measure digitalization disclosure. Prior studies found that digitalization disclosure in various platforms, such as websites, integrated reports, announcements, statutory filings, conference calls and earnings forecasts, had a positive effect on share price (Chen & Srinivasan, 2019; Hossnofsky & Junge, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021) and firm performance (Zeng et al., 2022). One study found that digitalization disclosure in annual reports reduced the risk of a share price crash (Jiang et al., 2022).

2.3. Hypothesis development

Artificial intelligence is rapidly changing industries and will disrupt how companies create, destroy and preserve value, resulting in new or modified business models (Pereira & Romero, 2017; Piccarozzi et al., 2018; Rayna & Striukova, 2016). Prior studies found that technologies driven by artificial intelligence positively impact value-creation processes. This value-creation process is achieved by creating new revenue streams, gaining access to a broader range of customers and suppliers, decreasing transaction time, optimization of production processes, lower production costs, and creating a more skilled workforce (Arnold et al., 2016; Nowiński & Kozma, 2017; Rachinger et al., 2019). Prior studies concluded that companies with higher digital activity had a higher firm value, and those companies that adopted blockchain technology positively affected a company's market value in the long term (Chen & Srinivasan, 2019; Hossnofsky & Junge, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021). Hossnofsky and Junge (2019) predicted that analysts' initial reactions to companies' digitalization efforts are adverse in the early years but became positive in later years. Therefore, it can be expected that the scope of digitalization disclosure positively impacts firm value. Therefore, the hypothesis is:

H_1*: The scope of digitalization disclosure captured in integrated reports has a positive impact on firm value.*

3. RESEARCH METHOD

The value relevance of digitalization disclosure is evaluated by executing the following steps. Firstly, a disclosure index to measure digitalization disclosure is developed and incorporated into a new text analysis software called the Fourth Industrial Revolution Disclosure Analysis Tool (4IRDAT), which uses algorithms based on natural language processing techniques to assess the scope of digitalization disclosure in integrated reports. Secondly, the results generated from 4IRDAT are interpreted using descriptive statistics to assess whether the scope of digitalization disclosure improved. Finally, the scores generated from 4IRDAT are inserted into the Ohlson (1995) Model to evaluate if digitalization disclosure affected share price, and panel regression is applied to generate the results.

3.1. Sample and data collection

Table 1 depicts the sample selection. The sample selection starts with the top one hundred JSE-listed companies based on market capitalization. The top one hundred companies represent the highest-performing companies on the JSE. Companies were eliminated from the sample due to dual listings, incomplete listings from 2018 to 2020, trading two types of shares, and limited available financial information. The final sample for the first dataset, which includes loss-making companies, is 90 companies (270 observations) and ranges for three years from 2018 to 2020. Eighteen companies incurred losses in the three years and were eliminated. The final sample for the second dataset, excludes loss-making companies, is 72 companies (216 observations). The sample is limited to three years because the technologies emerging in the 4IR may only be disclosed in integrated reports more recently.

	No. of companies	**No. of observations**
Sample size	100	300
Incomplete listings in 2018, 2019, 2020	(3)	(9)
Companies with limited financial info on IRESS	(5)	(15)
Dual listed companies	(1)	(3)
Companies trading two types of shares	(1)	(3)
Sample – including loss-making companies	**90**	**270**
Loss-making companies	(18)	(54)
Sample – excluding loss-making companies	**72**	**216**

Table 1. Sample

Information about digitalization disclosure was obtained from companies' integrated reports, and the process to measure digitalization disclosure is detailed in section 4.3. The integrated reports were downloaded in PDF from Equity RT or the respective company's websites. In cases where some reports were interactive or encrypted, and information could not be copied and pasted, optical character recognition (OCR) software was used to extract the information from the integrated reports into a readable format.

3.2. Model specification

The Ohlson (1995) Model was chosen to evaluate the value relevance of the scope of digitalization disclosure. The Ohlson (1995) Model is an accounting-based valuation model in which the market value of equity is a function of book value, accounting earnings, and other financial and non-financial value-relevant information and is derived as follows (Ohlson, 1995):

$$P_{i,t} = \propto_0 bv_{i,t} + \propto_1 x^a_{i,t} + \propto_2 v^a_{i,t} + e_{i,t} \qquad (1)$$

where:

$P_{i,t}$ = market value of equity in the current period t,

$bv_{i,t}$ = closing book value of equity for the period t,

$x^a_{i,t}$ = abnormal earnings for the period t,

$v_{i,t}$ = other information,

$e_{i,t}$ = standard error term.

Abnormal earnings were calculated using net income less a charge for capital. The charge for capital was calculated as the opening book value multiplied by the weighted average cost of capital (Swartz, 2006). Abnormal earnings were calculated as follows:

$$x^a_{i,t} = x_{i,t} - \text{WACC}\left(bv_{i,t}\right) \qquad (2)$$

where:

$bv_{i,t}$ = opening book value for the period t,

$x_{i,t}$ = current net income for the period t,

WACC = percentage obtained from IRESS.

A new non-financial variable was introduced into the Ohlson (1995) Model for the scope of digitalization technologies, named $4IRQ_i$. The process of measuring $4IRQ$ is detailed in section 4.3. Equation (1) was modified to include the new variable:

$$P_{i,t} = \propto_0 bv_{i,t} + \propto_1 x^a_{i,t} + \propto_2 4IRQ_{i,t} + \delta_i + \delta_t + \mu_{i,t} \qquad (3)$$

where:

$P_{i,t}$ = market value per share three months after year end,

$bv_{i,t}$ = closing book value per share for the period t,

$x_{i,t}^{a}$ = abnormal earnings for the period t, calculated as per Equation (2),

$4IRQ_{i,t}$ = the scope of digitalization disclosure is detailed in section 4.3,

δ_i = fixed cross-sectional effects,

δ_t = period effects,

$\mu_{i,t}$ = an error term.

The Ohlson (1995) Model was run using panel data in EViews, a statistical software. The Redundancy Fixed Effect-Likelihood Ratio and Hausman Test were applied to decide which panel data estimation technique was more appropriate to the different data sets. The Redundancy Fixed Effect-Likelihood Ratio evaluates whether cross-section and period fixed effects should be included in the model. A prior study suggested using the Hausman Test to choose between a fixed-effects model and a random-effects model (Onali et al., 2017).

3.3. Measurement of non-financial variable ($4IRQ_i$)

Content analysis has been characterized as "a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit coding rules" (Stemler, 2001). This research technique allows the researcher to "make replicable and valid inferences from texts, or other meaningful matter to the contexts of their use" (Krippendorff, 2004, p. 18). It also facilitates the identification and analysis of trends and patterns in large volumes of information by defining rules of coding (Stemler, 2001; Weber, 1990). Krippendorf (2004) discusses using a computer to aid content analysis and concludes that the software should provide information that can be interpreted to answer the research questions. In this paper, the 4IRDAT software has been developed to assess the scope of digitalization disclosure in integrated reports.

The steps set out by Weber (1990) for content analysis, were applied in this paper. Firstly, the recording units were defined as all text in the integrated reports. Secondly, the assessed categories were defined as the words in the digitalization dictionary. The third step was to define the coding rules for the scope of digitalization disclosure. A scoring index was developed to score digitalization disclosure: the final score was called 4IR quantity, abbreviated to $4IRQ_i$. The fourth step was to manually evaluate the integrated reports of twenty companies using the digitalization dictionary and scoring criteria. The results were recorded on an Excel spreadsheet. The keywords present were counted and inserted into the Excel spreadsheet. The application of the disclosure index for the scope of digitalization disclosure was conducted manually by searching for words using the control F function and manually scoring the disclosure based on the scoring index tables. The formulae to calculate the digitalization scores were also inserted into Excel. Further words were identified from this process and added to the digitalization dictionary. Thereafter, in step five, the software developers designed 4IRDAT, which is discussed in section 4.4. The software was tested, the scores were compared to the manual results, and the final version of 4IRDAT was developed. Finally, the sample of 270 integrated reports was scored using the 4IRDAT software. The scores were generated on 4IRDAT and exported to Excel. For further analysis, the individual company scores were merged into one spreadsheet for all three years.

Keywords relating to digitalization were identified from the digital dictionary developed by Hossnofsky and Junge (2019). The original digital dictionary comprised 54 keywords (Hossnofsky & Junge, 2019). Additional keywords were identified from previous studies and the initial manual assessment of twenty integrated reports, which resulted in an additional 32 keywords being added to the digitalization dictionary. The digitalization dictionary in Hossnofsky and Junge (2019) did not include more recent technologies. Table 2 below provides the list of keywords in the digitalization dictionary.

3D printing	Cloud technology	Digitalization	Natural language processing
4IR*	Cobots	Digitization	Neural network
Additive manufacturing	Connected car	E-business	New economy
AI*	Connectivity	E-catalogue	Newsfeed
Algorithms	Cryptocurrency	E-commerce	Online platform
App*	Cyber physical	E-learning	Open source code sharing
Artificial intelligence	Dark factory	E-mobility	Robotics
Augmented reality	Data analytics	E-procurement	Self-driving car
Autonomous technology	Data architecture	E-publishing	Sharing economy
Big data	Data capture	E-service	Smart content
Big data analytics	Data integration	E-travel	Smart devices
Biometric	Data mining	Fintech	Smart factory
Bitcoin	Data monetization	Fourth Industrial Revolution	Smart home

Blockchain	Data processing system	FutureSmart Mining	Smart phone
Bots	Data science	High-tech	Social media
Business intelligence	Deep learning	Industry 4.0	Software
Chatbot	DigiMine	Influencer	Software as a service
Chief information officer	Digital	Information architecture	Technology
Click through rate	Digital assets	Internet	Virtual reality
Cloud	Digital financial services	Internet of things	Web based
Cloud computing	Digital twin	IoT*	
Cloud kitchens	Digital wallet	Machine learning	

*Abbreviations considered more widely used

Table 2. Digitalization dictionary

Table 3 is used to generate a score for each company. When the company does not disclose any keywords from the digitalization dictionary, a score of zero is awarded. A score of 1 is awarded if the company discloses less than ten keywords from the digitalization dictionary. In cases where the company disclosed between ten and fourteen of the keywords from the digitalization dictionary, a score of 2 is awarded. A score of 3 is awarded when the company discloses between fifteen and twenty keywords from the digitalization dictionary. In cases where the company disclosed more than twenty of the keywords from the digitalization dictionary, a score of 4 is awarded. The final score was calculated by multiplying the score awarded by the maximum possible score.

Score	**Criteria**
0	No disclosure of any of the keywords from the digitalization dictionary.
1	Disclosed less than ten keywords from the digitalization dictionary.
2	Disclosed between ten and fourteen keywords from the digitalization dictionary.
3	Disclosed between fifteen and twenty keywords from the digitalization dictionary.
4	Disclosed more than twenty keywords from the digitalization dictionary.

Table 3. Scoring criteria for digitalization disclosure

$4IRQ_i$ is the final score generated from 4IRDAT, calculated as a percentage of the words present in integrated reports based on the scoring criteria in Table 3. $4IRQ_i$ is calculated as:

$$4IRQ_i = \frac{Quality\ score\ based\ on\ scoring\ index\ table}{Maximum\ quality\ score} \qquad (4)$$

where:

$4IRQ_i$ = Fourth Industrial Revolution quantity index

Maximum quality score is the highest score in the scoring criteria in Table 3, which is 4.

3.4. Development of 4IRDAT software

A software named 4IRDAT was developed to calculate $4IRQ_i$. 4IRDAT was developed in Microsoft's Visual Studio using #C programing language. Regular expression libraries were used for text processing. The criteria and calculations discussed in section 4.3 were programmed into the software. 4IRDAT uses algorithms based on natural language processing techniques to facilitate the content analysis of digitalization disclosure. Using 4IRDAT allowed textual data from integrated reports to be processed reliably and removed the bias of manually evaluating large amounts of text. 4IRDAT software facilitated the search for character strings, texts, and computation of scores for the assessment of $4IRQ_i$. The software and its functions are discussed next.

Figure 1 is the home screen and prompts the user to select the name of the company and the year being evaluated.

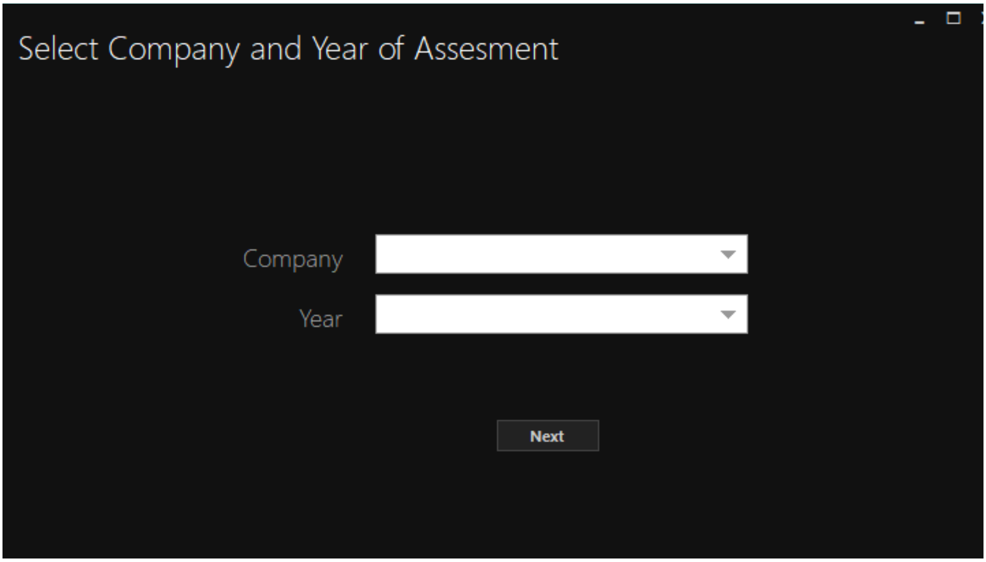

Figure 1. Screenshot of 4IRDAT – Select company and year

The screen in Figure 2 allows the user with the option to paste the text from the integrated report in the textbox "Report text". The entire integrated report was copied and pasted into this box.

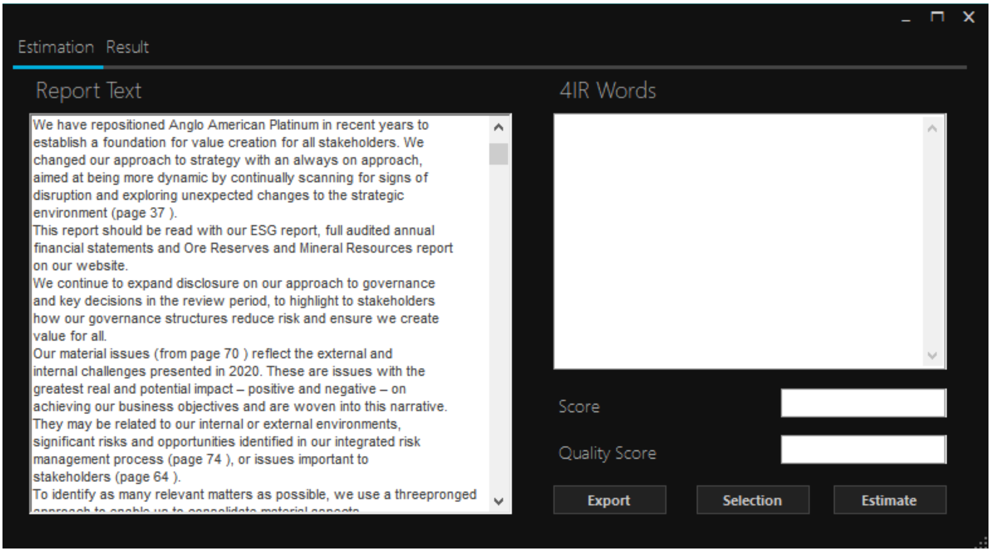

Figure 2. Screenshot of 4IRDAT – Paste integrated report

The score is displayed on the quality score textbox in Figure 3. Once the user clicks on "Estimate", the software identifies the words from Table 2 and generates a score for $4IRQ$ based on the scoring criteria in Table 3. The software makes use of the namespace system.Text.RegularExpressions. The function wordComparision with its associated parameters are designed to compare and count words or phrases that are similar in the report text, to the words in the keyword repository from Table 2. The report text is stripped of special characters using Unicode as a pattern in Regex. The report text is stripped of carriage return characters, and hyphens are replaced with spaces to ensure the string is cleaned. The keywords relating to digitalization are used as a pattern between word-boundaries. A counter variable is initialized to 0, and each loop is used to iterate through each keyword, and each keyword is trimmed off the newline character. An if-statement is with the condition that the Regex matches the keyword with a word in the report text. If true, the counter is increased by 1 and the keyword is displayed on the 4IR words textbox. The function returns the counter result. The function is called, and the result is used in the if-statement to determine the report's score, according to Table 3. The score is displayed in the score textbox. The quality score is calculated by finding the product of $4IRQ$ formula and 100, which is stated in equation (3).

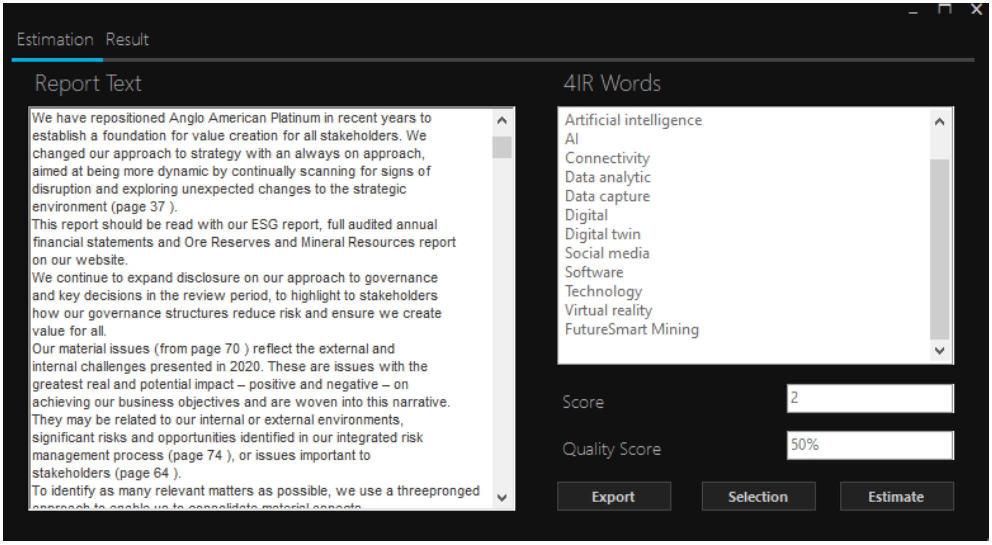

Figure 3. Screenshot of 4IRDAT – $4IRQ_i$ score

4. EMPIRICAL ANALYSIS AND RESULTS

The spread of scores and word frequency analysis are presented, followed by descriptive statistics, correlation analysis, and panel regression results.

4.1. Analysis of the spread of scores

An analysis of the spread of $4IRQ_i$ is presented in Table 4 below. The analysis indicated that none of the companies in the sample scored 0 in any of the three years. A score of 1 was obtained by 61 companies in 2018, 51 companies in 2019 and 49 companies in 2020. Most of the sample obtained a score of 1, indicating that most companies disclosed less than ten keywords from the digitalization dictionary. The number of companies that scored 1 decreased from 2018 to 2020, indicating a shift towards using more digitalization words in integrated reports. Twenty-two companies obtained a score of 2 in 2018, 26 companies in 2019 and 27 companies in 2020. Those companies that scored 2 disclosed 10-14 keywords from the digitalization dictionary. 6 companies obtained a score of 3 in 2018, 11 in 2019 and 10 in 2020. A score of 4 was obtained by 5 companies in 2018, 6 in 2019 and 8 in 2020. Few companies disclosed more than 20 keywords from the digitalization dictionary. These results indicated a slow, deliberate shift towards digitalization in South African listed companies.

Scores	Years		
	2018	2019	2020
0	0	0	0
1	61	51	49
2	22	26	27
3	6	11	10
4	5	6	8

Table 4. Analysis of the spread of $4IRQ_i$ scores across companies

4.2. Word frequency

Figure 4 depicts the trend of digitalization words disclosed from 2018 to 2020. There was an increase in the disclosure of digitalization words from 2018 to 2020. Cumulatively, the words from the digitalization dictionary were mentioned 2731 times in the total sample between 2018 and 2020. The words with the highest disclosure were identified from the results. The highest twenty words disclosed were technology, digital, software, social media, app, internet, cloud, connectivity, data analytic, artificial intelligence, AI, e-commerce, robotic, e-learning, online platform, fintech, machine learning, business intelligence, chief information officer and big data. An analysis of the highest twenty words revealed that although some words increased between 2018 and 2020, words such as software, AI, business intelligence, chief information officer and big data showed a decreasing trend. These words could be due to the use of abbreviations in reports such as AI (artificial intelligence), BI (business intelligence) and CIO (chief information officer). In addition, software and big data could have decreased because companies may be disclosing specific types of software and big data instead of the words "software" and "big data".

The words with the highest mentions were technology, digital and software. The word "technology" was mentioned a cumulative total of 282 times in the sample over the three years, which indicated that companies were enhancing their business processes using different types of technology. Notably, "digital" had 232 mentions, and "software" was mentioned 192 times cumulatively in the three years. The word "AI" had 87 mentions. The word "Fourth Industrial Revolution" appeared in 49% of the integrated reports in the sample.

Words with less than ten mentions included Industry 4.0, data architecture, digital twin, cryptocurrency, cloud technology, connected car, information architecture, 3D printing, deep learning, natural language processing, digitization, digitalization and smart factory. Eighteen words from the digitalization dictionary were not disclosed in any reports. These words included dark factories, cobot, cyber-physical, and neural networks, considered emerging technologies. These results indicate an increasing trend towards digitalization over the three years, and companies in South Africa are starting to embrace digitalization in their businesses. These findings are consistent with the findings made by Chen and Srinivasan (2019), in which they also noted an increase in the disclosure of digital activities, with the highest words being "analytics" and "digitization". However, in the sample of South African listed companies, analytics and digitization were not identified as the highest words.

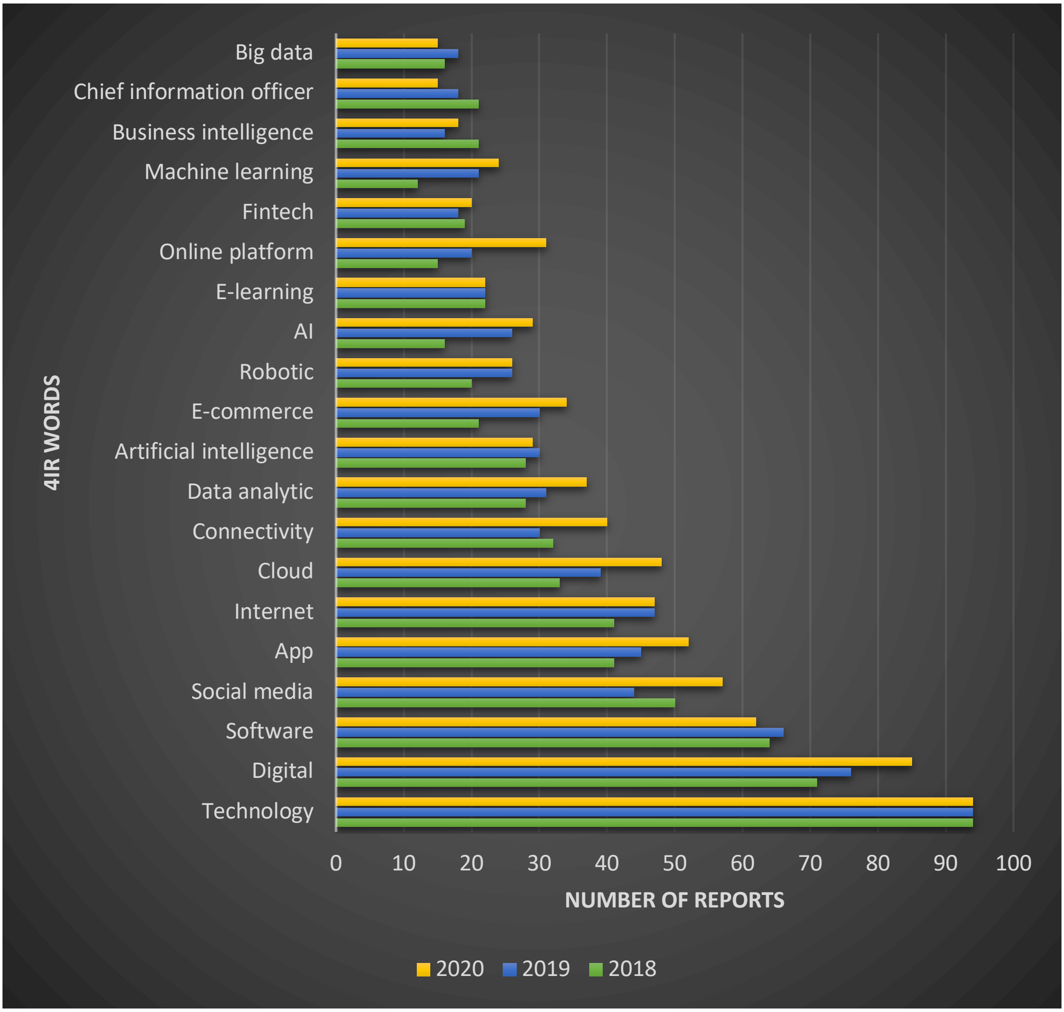

Figure 4. Trend analysis of digitalization words

4.3. Descriptive statistics and correlation analysis

Table 5 provides the descriptive statistics for the variables utilized in the panel regression for the samples, including loss-making companies and excluding loss-making companies.

Scope of digitalization disclosure						
Including loss-making companies						
Variable	n	**Mean**	**Median**	**Std dev**	**Min**	**Max**
$\boldsymbol{P}$	270	144.119	57.440	391.716	1	3485.000
$\boldsymbol{bv}$	270	61.597	31.265	103.893	1.1	965.410
$\boldsymbol{x^a}$	270	12.377	1.490	83.574	-28.390	875.650
$\boldsymbol{4IRQ}$	270	41.481	25.000	22.745	25.000	100.000
Excluding loss-making companies						
Variable	n	**Mean**	**Median**	**Std dev**	**Min**	**Max**
$\boldsymbol{P}$	216	162.710	62.525	434.111	1.000	3485.000
$\boldsymbol{bv}$	216	62.171	29.330	110.530	1.100	965.410
$\boldsymbol{x^a}$	216	16.138	2.165	93.011	-12.200	875.650
$\boldsymbol{4IRQ}$	216	42.361	25.00	23.243	25.000	100.00

Note: P, bv and x^a are in South African Rands. 4IRQ is in percentages.

Table 5. Descriptive statistics – Scope of digitalization disclosure

The results for the sample, including loss-making companies, are discussed first, followed by the results for the sample excluding loss-making companies.

Including loss-making companies

The average company in the sample had a market value of R144.12 per share. The median is R57.440. The standard deviation for market value is R391.72 per share. The minimum market value per share is R1 and the maximum is R875.65. The average company in the sample had a book value of R61.60 per share. The median book value is R31.26. The standard deviation for book value is R103.89. The minimum book value per share is R1.10 per share and the maximum is R965.41. The average company in the sample had abnormal earnings of R12.38 per share. The median is R1.49 and the standard deviation is R83.57. The minimum abnormal earnings per share is -R28.39 and the maximum is R83.57. The negative minimum abnormal earnings is due to some companies making losses, resulting in negative abnormal earnings. The average score for $4IRQ$ is 41.48%. The median is 25% and the standard deviation for $4IRQ$ is 75%. The minimum $4IRQ$ score is 25% and the maximum is 100%.

Excluding loss-making companies

The average company in the sample had a market value of R162.71 per share. The median is R62.53. The standard deviation for market value is R434.11 per share. The minimum market value per share is R1 and the maximum is R3 485. The average company in the sample had a book value of R62.17 per share. The median book value is R29.33. The standard deviation for book value is R110.53. The minimum book value per share is R1.10 per share, and the maximum is R965.41. The average company in the sample had abnormal earnings of R16.14 per share. The median is R2.17 and the standard deviation is R93.01. The minimum abnormal earnings per share is -R12.20 and the maximum is R875.65. The negative minimum abnormal earnings is due to some companies making losses which results in negative abnormal earnings. The average score for $4IRQ$ is 42.36% The median is 25% and the standard deviation for $4IRQ$ is 23.24%. The minimum $4IRQ$ score is 25% and the maximum is 100%.

Pearson Correlation

The results of the Pearson Correlation between market value, book value, abnormal earnings and $4IRQ$ are provided in Table 6. The results indicate that book value and abnormal earnings positively and significantly correlated with market value for both samples, including and excluding loss-making companies. In contrast, there is a weak but positive significant correlation between market value and $4IRQ$ for both samples.

	P	
	Including loss-making companies	**Excluding loss-making companies**
Variables	**Pearson (p-value)**	**Pearson (p-value)**
bv	0.898 (0.000)***	0.925 (0.000)***
x^a	0.911 (0.000)***	0.915 (0.000)***
$4IRQ$	0.169 (0.006)***	0.183 (0.007)***

KEY: *** Significant at the 0.01 level

Table 6. Pearson correlations between market value and book value, abnormal earnings and 4IRQ

4.4. Regression results

Table 7 provides the results for the value relevance of $4IRQ$. Table 8 provides the results for the Redundant Fixed Effect-Likelihood Ratio and the Hausman Test.

Model: $P_{i,t} = \propto_0 bv_{i,t} + \propto_1 x^a_{i,t} + \propto_2 4IRQ_{i,t} + \delta_i + \delta_t + \mu_{i,t}$

Variables	Including loss-making companies	Excluding loss-making companies
Constant	12.362 (0.585)	2.301 (0.934)
$\boldsymbol{bv_{i,t}}$	1.689 (0.000)***	2.063 (0.000)***
$\boldsymbol{x^a_{i,t}}$	2.461 (0.000)***	2.041 (0.000)***
$\boldsymbol{4IRQ}$	-0.065 (0.870)	-0.019 (0.969)
Period effects	No	No
Cross section effects	Random	Random
$\boldsymbol{n}$	270	216
Adj R^2	0.777	0.788

KEY: *** Significant at the 0.01 level.

Table 7. Regression results – 4IRQ

		Including loss-making companies	Excluding loss-making companies
Redundant Fixed Effect-Likelihood Ratio	Cross-section F	0.000	0.000
	Cross-section Chi-squared	0.000	0.000
	Period F	0.288	0.452
	Period Chi-square	0.147	0.291
Hausman Test	Cross-section random	0.971	0.864
	Period random	-	-

Table 8. Redundant Fixed Effect-Likelihood Ration and Hausman test – 4IRQ

The results for the sample, including loss-making companies, are discussed first, followed by the results for the sample excluding loss-making companies.

Including loss-making companies

The most appropriate model for the dataset, including loss-making companies is the random cross-section effects model, which has a high explanatory power of 78%.

The co-efficient for $bv_{i,t}$ and $x_{i,t}^{a}$ are positive and significant at the 1% level. The co-efficient for $4IRQ$ is -0.065 with a p-value of 0.870. The co-efficient for $4IRQ$ is negative and insignificant. These results indicate a negative, but insignificant relationship between $P_{i,t}$ and $4IRQ$, which mean that $4IRQ$ has no effect on share price when loss-making companies are included in the sample.

The scores for the Redundant Fixed Effect-Likelihood Ratio are presented in Table 7. The Chi-square for cross-section is 0.000; therefore, the null hypothesis that fixed effects are redundant is rejected, and heterogeneity must be accounted for in the cross-sections. The Chi-square for period effects is 0.147; therefore, the null hypothesis that period effects are redundant is accepted, and heterogeneity is not accounted for in the period effects. The scores for the Hausman Test are presented in Table 7. The Chi-square for cross-section effects is 0.971; therefore, the null hypothesis that the random cross-section effects model is the best model is accepted, and random cross-section effects are the final model.

Excluding loss-making companies

The most appropriate model for the dataset, excluding loss-making companies is the random cross-section effects model, which has a high explanatory power of 79%. The co-efficient for $bv_{i,t}$ and $x_{i,t}^{a}$ are positive and significant at the 1% level. The co-efficient for $4IRQ$ is -0.019 with a p-value of 0.969. These results indicate a negative, but insignificant relationship between $P_{i,t}$ and $4IRQ$. Therefore, $4IRQ$ has no effect on share price when loss-making companies are eliminated from the sample.

The scores for the Redundant Fixed Effect-Likelihood Ratio are presented in Table 7. The Chi-square for cross-section is 0.000; therefore, the null hypothesis that fixed-effects are redundant is rejected, and heterogeneity must be accounted for in the cross-sections. The Chi-square for period effects is 0.291; therefore, the null hypothesis that period effects are redundant is accepted, and heterogeneity is not accounted for in the period effects. The scores for the Hausman Test are presented in Table 7. The Chi-square for cross-section effects is 0.864; therefore, the null hypothesis that the random cross-section effects model is the best model is accepted, and random cross-section effects is selected as the final model.

H_1 predicted that the incorporation of digitalization disclosure in integrated reports has a positive impact on firm value. Both models have a negative but insignificant relationship between $P_{i,t}$ and $4IRQ$, thus H_1 is rejected for both scenarios. The research question in this paper is - Is the scope of digitalization disclosure in integrated reports value-relevant for investors? The results indicate that digitalization disclosure has no effect on share price and is not value-relevant. No prior studies have established the value-relevance of digitalization disclosure in integrated reports for South African companies.

In contrast to the findings in this paper, prior studies found that digitalization disclosure is value-relevant (Chen & Srinivasan, 2019; Lui & Ngai, 2019; Ricci et al., 2020; Salvi et al., 2021). Hossnofsky and Junge (2019) found that analysts react negatively to initial digital transformation, and in later years, the reaction becomes less unfavorable. The same concept could be applied to investors' responses to digitalization disclosure in the integrated reports of South African-listed companies. This paper concludes that the South African market has not reacted to digitalization disclosure in integrated reports. The practical implication is that the study illustrated that there is limited research on the value relevance of digitalization disclosure in companies in Africa. The level of digitalization and value relevance results for the sample companies indicate that further development is required for Africa to attain the objectives outlined in the African Union's Digital Transformation Strategy for Africa (2020 to 2030).

5. CONCLUSIONS AND DISCUSSION

This paper evaluated the value relevance of digitalization disclosure of the top 100 South African-listed companies. This paper develops a customized disclosure index using algorithms based on natural language processing techniques to evaluate digitalization disclosure in integrated reports. It provides a platform for developing customized software to enable the content analysis of unstructured data in financial and non-financial reporting. This paper contributes to the literature on the measurement and value relevance of digitalization disclosure in integrated reports. The results provide insights into the level of digital transformation in a sample of South African listed companies. The study illustrated highlights that there is limited research on the value relevance of digitalization disclosure in companies in Africa. The level of digitalization and the value relevance results for the companies in the sample indicate there is more development necessary in order for Africa to achieve the goals set out in the African Union's Digital Transformation Strategy for Africa (2020 to 2030).

An increase in digitalization disclosure over the three years was noted, with the words "technology," "digital", and "software" as the highest mentioned words from the digitalization dictionary. These findings are consistent with those of Chen and Srinivasan (2019), who found an increased disclosure of digital activities. Although most companies in the sample disclosed the word "technology," more detailed disclosure is required on the specific technologies applied in their businesses as the 4IR progresses. Similarly, more in-depth disclosure is required of the aspects related to "digital" and "software".

The regression results indicated that digitalization disclosure had yet to be incorporated into the share price of South African companies. Mapingire et al., (2021) highlight that South African companies have begun implementing digital transformation strategies within the last ten years. Therefore, it will take time to fully implement and reflect in their corporate reporting and share prices. These findings contrast with prior studies by Chen and Srinivasan (2019), Hossnofsky and Junge (2019), Ricci et al. (2020) and Salvi et al. (2021), which found that digitalization disclosure had a positive effect on the share price.

This paper has a few limitations. The measure for the scope of digitalization disclosure was limited to the presence of specific words due to the limited disclosure of digitalization available in integrated reports. The content analysis method in 4IRDAT, which relies on identifying keywords, may not necessarily reflect the context in which the word was discussed. Furthermore, there is no reporting framework or guideline for preparers on digitalization disclosure. The disclosure of digitalization in integrated reports and actual digitalization activity may differ. However, one would expect companies to disclose major digitalization projects and activities in their integrated reports to benefit their stakeholders. Investors use many sources of information to make decisions, and the integrated report is one of the sources. The integrated reports were used as the primary source of information in this paper. The results for the value relevance of the scope of digitalization disclosure should be interpreted cautiously because the results may be influenced by other factors not examined in this paper. The R^2 of the final regression models selected were more than 70% and considered a good fit. The results in this paper are consistent and can be replicated.

The 4IR provides many opportunities for future research. Firstly, 4IRDAT could be enhanced with machine learning techniques, and the 4IR dictionary could be further developed to include more industry-specific technologies and the latest developments in 4IR technologies. Secondly, evaluating and comparing digitalization disclosure in developed and developing countries could be useful in determining the progress of 4IR globally. Thirdly, exploratory research on the evolutionary process of successfully transforming to digital business models could provide insights into the transition. Fourthly, digitalization disclosure on the company's web pages and social media could be evaluated to assess the progression of the transition to the 4IR by using advanced algorithms which continuously identify and assess disclosure on social media and the internet. More advanced econometric models could be developed to assess the value relevance of digitalization disclosure and applied to larger samples across time, industries, and countries.

7. REFERENCES

Arnold, C., Kiel, D., & Voigt, K. I. (2016). How the industrial internet of things changes business models in different manufacturing industries. *International Journal of Innovation Management*, *20*(8), 1–20. https://doi.org/10.1142/S1363919616400156

Chen, W., & Srinivasan, S. (2019). Going Digital: Implications for Firm Value and Performance. *Harvard Business School Working Paper*, 19–117. https://dx.doi.org/10.2139/ssrn.4177947

Chowdhary, K. R. (2020). Fundamentals of artificial intelligence. In *Fundamentals of Artificial Intelligence*. Springer Nature India. https://doi.org/https://doi.org/10.1007/978-81-322-3972-7

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245. https://doi.org/10.1016/j.jacceco.2017.07.002

El-Haj, M., Rayson, P., Walker, M., Young, S., & Simaki, V. (2019). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance and Accounting*, *46*(3–4), 265–306. https://doi.org/10.1111/ jbfa. 12378

El-Haj, M., Alves, P., Rayson, P., Walker, M., & Young, S. (2020). Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as PDF files. *Accounting and Business Research*, *50*(1), 6–34. https://doi.org/10.1080/ 00014788.2019.1609346

Fama, E., & French, K. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance*, *LI*(1). https://doi.org/10.1111/j.1540-6261.1996.tb05202.x

Hossnofsky, V., & Junge, S. (2019). Does the market reward digitalization efforts? Evidence from securities analysts' investment recommendations. *Journal of Business Economics*, *89*(2019), 965–994. https://doi.org/10.1007/s11573-019-00949-y

Ibarra, D., Ganzarain, J., & Igartua, J. I. (2018). Business model innovation through Industry 4.0: A review. *Proceedings of the 11th International Conference Interdisciplinarity in Engineering*, *October 5-6*, 1–7. https://doi.org/10.1016/j.promfg.2018.03.002

IoDSA. (2016). *King IV Report on Corporate Governance*. IoDSA. https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/684B68A7-B768-465C-8214-E3A007F15A5A/IoDSA_King_IV_Report_-_WebVersion.pdf Accessed on 1 August 2023

Jiang, K., Du, X., & Chen, Z. (2022). Firms' digitalization and stock price crash risk. *International Review of Financial Analysis*, *82*, 102196. https://doi.org/10.1016/j.irfa.2022.102196

Kao, A., & Poteet, S. (2007). *Natural Language Processing and Text Mining*. Springer London. https://doi.org/10.1007/978-1-84628-754-1

Klingenberg, C. O., Borges, M. A. V., & Antunes, J. A. V. (2021). Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. *Journal of Manufacturing Technology Management*, *32*(3), 570–592. https://doi.org/10.1108/JMTM-09-2018-0325

Köseoğlu, S. D. (2022). *Financial Data Analytics*. Springer International Publishing. https://doi.org/10.1007/978-3-030-83799-0

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (Second). Sage Publications.

Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, *49*(5), 587–615. https://doi.org/10.1080/00014788.2019.1611730

Li, G., Hou, Y., & Wu, A. (2017). Fourth Industrial Revolution: technological drivers, impacts and coping methods. *Chinese Geographical Science*, *27*(4), 626–637. https://doi.org/10.1007/s11769-017-0890-x

Lui, A. K. H., & Ngai, E. W. T. (2019). The long term effect of blockchain adoption on firm value. *Proceedings of the 25th Americas Conference on Information Systems*.

https://aisel.aisnet.org/amcis2019/strategic_uses_it/strategic_uses_it/5/ Accessed on 25 August 2023.

Mapingire, K., Smuts, H., & Van der Merwe, A. (2021). Components of a Digital Transformation Strategy: A South African Perspective. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Sixth International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, vol 236. Springer, Singapore. https://doi.org/10.1007/978-981-16-2380-6_38

Matthies, B., & Coners, A. (2015). Computer-Aided Text Analysis of Corporate Disclosures - Demonstration and Evaluation of Two Approaches. *The International Journal of Digital Accounting Research*, *15*, 69–98. https://doi.org/10.4192/1577-8517-v15_3

Miura, T., Furukawa, T., Harada, J., Hirano, Y., & Hashimoto, T. (2021). Evaluation of Universities' Integrated Reports Using Text Mining Technique. *2021 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2021*, 1–6. https://doi.org/10.1109/SOLI54607.2021.9672410

Nakagawa, K., Sashida, S., Kitajima, R., & Sakai, H. (2020). What Do Good Integrated Reports Tell Us?: An Empirical Study of Japanese Companies Using Text-Mining. *Proceedings - 2020 9th International Congress on Advanced Applied Informatics, IIAI-AAI 2020*, 516–521. https://doi.org/10.1109/IIAI-AAI50415.2020.00108

Nowiński, W., & Kozma, M. (2017). How can blockchain technology disrupt the existing business models? *Entrepreneurial Business and Economics Review*, *5*(3), 173–188. https://doi.org/10.15678/EBER.2017.050309

Ohlson, J. (1995). Earnings, book-values, and dividends in equity valuation. *Contemporary Accounting Research*, *11*(11), 661–687. https://doi.org/10.1111/j.1911-3846.1995.tb00461.x

Onali, E., Ginesti, G., & Vasilakis, C. (2017). How should we estimate value-relevance models? Insights from European data. *British Accounting Review*, *49*(5), 460–473. https://doi.org/10.1016/j.bar.2017.05.006

Pereira, A. C., & Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, *13*, 1206–1214. https://doi.org/10.1016/j.promfg.2017.09.032

Piccarozzi, M., Aquilani, B., & Gatti, C. (2018). Industry 4.0 in Management Studies: A Systematic Literature Review. *Sustainability*, *10*(10), 3821. https://doi.org/10.3390/su10103821

Rachinger, M., Rauter, R., Müller, C., Vorraber, W., & Schirgi, E. (2019). Digitalization and its influence on business model innovation. *Journal of Manufacturing Technology Management*, *30*(8), 1143–1160. https://doi.org/10.1108/JMTM-01-2018-0020

Rayna, T., & Striukova, L. (2016). 360° Business Model Innovation: Toward an Integrated View of Business Model Innovation. *Research-Technology Management*, *59*(3), 21–28. https://doi.org/10.1080/08956308.2016.1161401

Ricci, F., Scafarto, V., Ferri, S., & Tron, A. (2020). Value relevance of digitalization: The moderating role of corporate sustainability. An empirical study of Italian listed companies. *Journal of Cleaner Production*, *276*, 123282. https://doi.org/10.1016/j.jclepro.2020.123282

Salvi, A., Vitolla, F., Rubino, M., Giakoumelou, A., & Raimo, N. (2021). Online information on digitalisation processes and its impact on firm value. *Journal of Business Research*, *124*, 437–444. https://doi.org/10.1016/j.jbusres.2020.10.025

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation*, *7*(17). https://doi.org/10.7275/z6fm-2e34

Swartz, G. (2006). *An Empirical Examination of the Ohlson (1995) Valuation Model in South Africa. Masters (Commerce) Thesis*. University of Witwatersrand.

Türegün, N. (2019). Text Mining in Financial Information. *Current Analysis on Economics & Finance*, *1*(647), 18–26. http://mesfordpublisher.com/wp-content/uploads/2019/03/Text-Mining-in-Financial-Information.pdf Accessed on 23 July 2023

Wagire, A. A., Rathore, A. P. S., & Jain, R. (2019). Analysis and synthesis of Industry 4.0 research landscape. *Journal of Manufacturing Technology Management*, *31*(1), 31–51. https://doi.org/10.1108/JMTM-10-2018-0349

Weber, R. P. (1990). *Basic Content Analysis* (Second). Sage Publications.

World Economic Forum. (2019). *World Economic Forum: Global Risks Report 2019*. World Economic Forum. https://www.weforum.org/reports/the-global-risks-report-2019/ Accessed on 5 June 2023.

Zeng, H., Ran, H., Zhou, Q., Jin, Y., & Cheng, X. (2022). The financial effect of firm digitalization: Evidence from China. *Technological Forecasting and Social Change*, *183*(2022), 121951. https://doi.org/10.1016/j.techfore.2022.121951

Zhang, M. C., Stone, D. N., & Xie, H. (2019). Text Data Sources in Archival Accounting Research: Insights and Strategies for Accounting Systems' Scholars. *Journal of Information Systems*, *33*(1), 145–180. https://doi.org/10.2308/isys-51979