



Article

Online Sustainability Reporting and Firm Performance: Lessons Learned from Text Mining

Xue Ning 1, Dobin Yim 2,* and Jiban Khuntia 3

- Information Systems, Business School, University of Colorado Denver, Larimer St, Denver, CO 80202, USA; xue.ning@ucdenver.edu
- Information Systems, Sellinger School of Business, Loyola University Maryland, 4501 North Charles St, Baltimore, MD 21210, USA
- Business School, University of Colorado Denver, Larimer St, Denver, CO 80202, USA; jiban.khuntia@ucdenver.edu
- * Correspondence: dyim@loyola.edu; Tel.: +1-410-617-2396

Abstract: As a corporate social responsibility (CSR) initiative, firms are increasingly disclosing sustainability indicators on online platforms to attract stakeholders' interests. It is vital to understand what indicators reflect more on a firm's performance and valuations. This study focuses on deriving value-oriented business intelligence from the voluntary disclosure of sustainability reports. The analysis in this study involves a three-stage approach: (1) Latent Dirichlet allocation (LDA) based topic modeling algorithm to identify and summarize typical contents expressed in various documents, (2) firm's sustainability maturity modeled as a function of its strategic intent using a latent Markov model (LMM) to estimate the statistical significance and the extent of their relationships, and (3) empirical analysis using random effect linear and non-linear probit models to explore the impact of antecedents and firm performance consequences of three strategic intents. This study highlights using an advanced business analytics approach, specifically with latent Dirichlet allocation (LDA) topic modeling, to codify intangible knowledge embedded in annual sustainability reports to infer a firm's strategic intent behind voluntary disclosure. In addition, this study aims to analyze the influence of the firm's sustainability strategic intent on its financial performance. A secondary panel dataset consisting of information on 680 firms in 3 years was constructed by matching the text mined data with information from other sources. Results indicate that, on the one hand, while external stakeholder engagement is the primary motivation behind voluntary disclosure of sustainability reporting, firms are starting to engage internal stakeholders through workforce practices. On the other hand, internal employee-oriented intent has more influence on firm performance than external customer-oriented intent. This study demonstrates a toolset to index firms' sustainability indicators and evaluates firms' sustainability practice as an intangible asset and its impact on firms' financial performance.

Keywords: online sustainability reporting; voluntary disclosure; topic modeling; firm performance; analytics



Citation: Ning, X.; Yim, D.; Khuntia, J. Online Sustainability Reporting and Firm Performance: Lessons Learned from Text Mining. Sustainability 2021, 13, 1069. https://doi.org/10.3390/su13031069

Academic Editor: Evangelos Katsamakas Received: 25 November 2020 Accepted: 18 January 2021 Published: 21 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Corporate social responsibility (CSR) initiatives create a positive signal to stakeholders, including investors, customers, and employees [1]. This is credence for firms, with executives recognizing the importance of sustainability for underlying reputation, marketing, and strategy linked business imperatives. The surge in sustainability interests is driven by the growing percentage of intangible assets representing market value, overtaking physical and financial assets, and more so in uncertain conditions. For instance, Ocean Tomo reported that intangible assets made up 84% of the S&P 500 market value in 2015, and the study suggested that due to the recent COVID-19 situation, this rate is more than 90% in 2020 [2].

Sustainability **2021**, 13, 1069 2 of 15

Investors are demanding clarity in the intangible assets component of firm valuation. Several rating agencies such as Bloomberg, Thomson-Reuters, Dow Jones Sustainability Index have started including sustainability indicators as part of the reporting process [3] to rank a firm's value using sustainability disclosure performance. One index the financial services industry has come to some consensus around is ESG, a catch-all term covering environmental, social, and governance issues and includes material to firm performance, risk, profitability, and firm viability. The process of measuring the ESG score includes questionnaires, interviews, and analyses of company reports such as annual sustainability reports, to name a few. The emergence of socially responsible investing (SRI) and impact investing has made ESG information collection services valuable to financial analysts and investors in the marketplace. Thereby, leading to a systems perspective of sustainability data collection, reporting, and subsequent outcomes derived from the impact on stakeholders as an essential value proposition for firms.

Systems representation of sustainability reporting emphasizes the underlying concept of information collection and dissemination—that is the sole purpose of information systems used for sustainability management, emerging as a vital research area in the information systems and related disciplines. In this domain, studies on sustainability often focused on IT adoption, use, and impacts on environmental sustainability, including carbon emission, energy consumption, and waste management [4,5]. The primary focus of the extant IS research was to evaluate traditional IT artifacts in a sustainability context. However, nowadays, online digital platforms have become a new data source for sustainability research [6].

Extending the prior research around information systems and sustainability, this study proposes an integrative view of information systems as a social responsibility platform (SRP), a digital ecosystem in which various organizations and stakeholders with a shared interest in developing and implementing CSR practices come to learn, adopt, and practice socially responsible behaviors. Information systems' primary role is to provide a digital platform composed of loosely connected information systems and service architecture to enable firms, non-governmental organizations (NGOs), and infomediaries to store, exchange, and use data on firms' CSR-related activities.

This study aims to utilize analytics techniques to extract concepts from firms' digital sustainability reports disclosed on a sustainability reporting platform. Our study's key focus is to provide insights into how to derive business intelligence from a knowledge repository of information artifacts that firms contribute to their voluntary disclosure of CSR related activities. A lack of accepted practices on reporting standards and metrics constrains the interpretability of an individual firm's CSR efforts into a comparative evaluation of firms across the industry. For instance, firms release annual sustainability reports that are text-based, often exceeding hundreds of pages, to shareholders, making it difficult to assess a firm's performance against past years, let alone compare against competitors. We introduce an advanced business analytics approach of codifying intangible knowledge embedded in annual sustainability reports to infer a firm's strategic intent behind voluntary disclosure. Then, we empirically test and validate the relationships between these strategic intents and a firm's performance.

We use a dataset obtained from the global reporting initiative (GRI)—a non-government organization that provides CSR reporting guidelines—that includes procedures and commonly used metrics to assess a firm's CSR-related performance. While GRI is not a certifying nor an enforcement body of CSR compliance, the reporting guidelines they provide are among the few guidelines that are gaining acceptance among organizations, especially those in the financial industry. We track 682 firms for the years 2013 to 2015 and analyze the sustainability reports.

The analysis is conducted using three steps. First, we use the latent Dirichlet allocation (LDA) based topic modeling algorithm to identify and summarize typical contents expressed in various documents. In the second stage, we model a firm's sustainability maturity as a function of its strategic intent using a latent Markov model (LMM) to estimate

Sustainability **2021**, 13, 1069 3 of 15

the statistical significance and the extent of their relationships. Furthermore, we ran a model with three hidden states and three topics as covariates to estimate the three topics influencing the hidden states' statistical significance in initial and transition probabilities. The third stage consists of an empirical analysis using random effect linear and non-linear probit models. The linear model links strategic intent to the firm's performance measured by the return on assets. The non-linear probit model estimates the relationship between three topics discussed in a firm's annual sustainability report and three types of firms' strategic intent on the environment, work, and customer.

Our findings indicate that firms' primary strategic intent to report practices is to manage a firm reputation to external stakeholders, such as customers in the marketplace. We also find that firms are starting to engage internal stakeholders such as employees through social discussion around labor practices. Furthermore, we find firms that report environmental performance as compliance reporting tends to stay committed. We also find the impacts of topics around the environment, social, and governance (ESG) on these three primary strategic intents' transition probabilities. We further find strongly significant and positive relationships between topics and strategic intents. Finally, we find that environment and employee-related strategic intents have significant and positive impacts on firm performance, while customer-related strategic intent does not significantly influence. This study demonstrates one approach to applying analytics for sustainability research and contributes to the research and practice by offering a nuanced view of the strategic intents embedded in firms' sustainability reports and their impacts on firm performance.

2. Related Literature

Sustainability is emerging as an important issue for firms in recent years. Therefore, corporate social responsibility, competitions, customer awareness, standards, and certification perspectives and requirements have led sustainability to be discussed at most board meetings, stakeholder assemblies, and policy forums [7]. Corporate social responsibility is nothing but firms' commitments to operating economically, socially, and environmentally sustainable [8]. A responsible firm aims to minimize harm and maximize benefits in its relationships with stakeholders. It is estimated that about one-fifth of assets under management in the United States and half of all assets under management in the European Union employ socially responsible investment strategies—but it is the other half that is highly concerning to the World as a whole [9].

Firms embrace corporate responsibility initiatives aligned with sustainability due to the effect on stakeholders such as investors, customers, and employees. Both practitioners and scholars recognize sustainability's strategic importance as one of the top strategic agendas [10]. Many chief executives feel that communicating the firm's sustainability activities to stakeholders is the top reason behind firm reputation management [11]. Previous studies have shown that only social responsibility disclosure, rather than environmental initiatives reporting, supports firm value [12].

When reported and disclosed to stakeholders, a firm's sustainability strategy presents a comprehensive view of its sustainability activities. Reliance on sustainability and responsibility performance ratings to assess historical and cross-industry comparability of firm value has become a norm recently [13]. These disclosures are influencing firm value perceived by market participants and attracting socially responsible investing.

Researchers have conducted numerous studies to explore the relationship between reported sustainability practice indicators and firm performance. Their findings are quite different, including positive, neutral, and negative relationships [14,15] and marginal, moderating, and controlling impacts with various variables [12]. Variables such as firm size, firm investment in research and developments, and industry type are suggested to be considered when studying the link between sustainability reporting and firm financial performance [15,16]. Despite a lack of direct impact of sustainability practices on firm performance, firms may have alternate motivations to undertake sustainability initiatives besides signaling firm quality to investors. One rationale for disclosing sustain-

Sustainability **2021**, 13, 1069 4 of 15

ability activities is managing relationships with external stakeholders through reputation management [17].

Prior research has shown that customers demand firms have sustainability as part of their corporate strategy [18,19] and show their support through brand loyalty [20]. Another rationale may be for managing reputation with internal stakeholders such as employee attraction and retention [21]. Firms leverage human capital as a core resource to operate and deliver products or services to customers. Promoting and maintaining a good reputation in terms of the right working environment can motivate a firm's workforce [22] and translate into a competitive edge against other firms [23]. Therefore, evaluating a firm's underlying intention and, more importantly, its commitment is crucial to understanding the functional role of sustainability disclosure for firms.

Prior studies have suggested that the indirect effect of non-financial indicators, such as that of sustainability, maybe through customer relations management [17], employee engagement such as attracting and retaining talent [21], or as a positive signal to stakeholders such as investors, customers, and employees [16]. Therefore, some firms benefit from the reporting process only when developing a strategy to collect, disclose, and take feedback on sustainability attributes.

Information systems scholars have recently given much attention to developing new theories on the role of information systems and technologies concerned with environmental sustainability issues relevant to industry, firms, communities, and individuals.

Previous literature applied various theories in sustainability, especially IS for Sustainability research. However, variants are also observed in sub-areas of studies, such as adoption, design, and implementations in relevant areas. We highlight the important ones in three groups. First, theories such as institutional theory, organizational theory, (natural) resource-based view (RBV), motivational theory, technology acceptance model (TAM), diffusion of innovation theory were used broadly in prior studies related to green IS/IT adoption. For example, institutional theory and organizational theory (sense-making perspective) were used to study IT-based environmental compliance management [24]. The natural resource-based theory was used to examine the relationship between executive compensation and IT-based environmental strategies execution [16,25]. Theories from motivational psychology were used to investigate the adoption of sustainable technologies [26].

Second, some theories, including the belief-action-outcome (BAO) framework, affordance theory, informedness theory, etc., were used to understand organizations' or individuals' green practices and green behaviors. For instance, the BAO framework was developed to explain the role of IS in shaping beliefs, enabling actions, and improving the outcome for environmental sustainability [27].

Third, theories in design science (e.g., real options theory) were used as means for solutions' creations and focus on solutions' impacts [4]. Some other theories used in sustainability research are actor-network theory [28], technology-organization-environment (TOE) framework [29], goal-setting theory [30], socio-technical perspective [31], absorptive capacity theory [32], agency theory [25], activity theory [33], and signaling theory [16].

The organizational theory and RBV can support the theoretical foundation for our research. On the one hand, as one increasingly crucial intangible asset, online sustainability reporting can create value for firm performance. On the other hand, factors such as environment motivation, employee, and customer-orientation are critical elements of organizational factors that can influence firm performance.

Much of green IS research has focused on reducing the environmental impact through efficient design or energy use [34], which considers the social impact on individuals, communities, wildlife, and the business impact on profitability. Second, a view of self-governance in which firms voluntarily act towards sustainable practices [27] is incomplete without extending beyond firms' operational boundaries. For instance, the theoretical model should incorporate external influences on a firm's decision to initiate sustainability practices. While niche markets such as impact investing or socially responsible investing suggest the viability of environmentally oriented firms, a lack of research on consensus

Sustainability **2021**, 13, 1069 5 of 15

building across organizations limit the generalizability of green IS research and calls for a more integrative view of environmental sustainability [8].

Through a review of the most recent literature, three broad areas of analytics for sustainability research are identified: (1) Leverage non-traditional and unstructured data related to sustainability, (2) online platform sustainability analytics, and (3) use advanced analytics for sustainability research.

First, using traditional IS can generate system-supported structured data. However, with the big data revolution, sustainability data can be found in social media such as Twitter, Facebook, and YouTube with different formats, including text, pictures, and video [35]. Researchers must leverage these different non-traditional and unstructured data and integrate them with traditional data sources for sustainability studies.

Second, the online platform provides another avenue to understand sustainability issues. For example, [6] studied the relationship between Internet-enabled matching platform and environmental benefits, and the study found that Craigslist's entry into a geographic market results in a 2–6% annual reduction in municipal solid waste per capita generated. Moreover, [36,37] studied the role of a competitive platform.

Third, various advanced analytics, such as mathematical modeling and machine learning techniques, have been applied in sustainability research [38]. All these insights gained from prior literature are reflected in our study. Sustainability data embedded in the sustainability reports are unstructured. We leverage such unstructured data from an online sustainability reporting platform and apply advanced analytics techniques in this study. By analyzing the unstructured data, we can better understand the firms' sustainability strategy and predict firms' performance.

3. Methods

The methods for this study involved three stages. In the first stage, we use the latent Dirichlet allocation (LDA) based topic modeling algorithm to identify and summarize typical contents expressed in various documents. In the second stage, we model a firm's sustainability maturity as a function of its strategic intent using a latent Markov model (LMM) to estimate the statistical significance and the extent of their relationships. We use the R programming library LMest to estimate the optimal number of hidden states that lead to our model's best fit. We further ran a model with three hidden states and three topics as covariates to estimate the three topics influencing the hidden states' statistical significance in initial and transition probabilities. In the third stage, random-effects linear and non-linear probit models are used for empirical analysis. The linear model links strategic intent to firm performance measured by the return on assets. The non-linear probit model estimates the relationship between three topics discussed in a firm's annual sustainability report and three types of firm's strategic intent on the environment, work, and customer. Due to the fact that the dependent variables are binary measures, we run the random effects probit model. We elaborate on our analysis of these three stages, along with the materials and process used, next.

3.1. Materials

A social responsibility platform (SRP) refers to a digital ecosystem in which various organizations and stakeholders with a shared interest in corporate social responsibility (CSR) come to learn, adopt, and practice socially responsible behaviors. The platform comprises loosely connected information systems that enable firms, NGOs, and infomediaries to store, exchange, and use data on firms' CSR activities. However, despite market interest in linking sustainability activities to firm valuation, a lack of compliance policy, standards, and infrastructure slows down the adoption and practice. The focus of this study is concerned with deriving business intelligence from the voluntary disclosure of sustainability reports. The study introduces an advanced business analytics approach, specifically with latent Dirichlet allocation (LDA) based on topic modeling, to codify intan-

Sustainability **2021**, 13, 1069 6 of 15

gible knowledge embedded in annual sustainability reports to infer a firm's strategic intent behind voluntary disclosure.

Our dataset comes from the global reporting initiatives (GRI) database. The vision of GRI reporting is to create "a sustainable global economy where organizations manage their economic, environmental, social and governance performance and impacts responsibly and report transparently." Its mission is to enable firms to make reporting standard practice by providing guidance and support to organizations. It is a toolset to make firm production activities transparent and accountable to internal and external stakeholders for their environmental impacts. The implementation a firm undertakes to disclose material topics related to sustainability reporting is choosing between (a) Core or (b) Comprehensive criteria, where the latter requires more extensive reporting of indicators according to the material aspect. There are six groups of indicators: (1) Economic, (2) labor, (3) product responsibility, (4) environment, (5) human rights, and (6) social. Based on a firm's maturity level, some or all of these indicators will be included in the report for the internal and external stakeholders to read, understand, and evaluate a firm's readiness to tackle the challenge of balancing a firm's interest with sustainability.

In other words, the GRI considers sustainability strategies to be reported that capture economic, environment, workforce (i.e., labor and employee), human rights, the customer (i.e., product responsibility), and societal perspectives of firm activities and provides measurements to monitor progress over time. Each indicator is a group of measurements which can be qualitative or quantitative. A firm can choose to report one or more activities, and what and how much to report can impact stakeholders' attitudes towards the firm. Most firms choose to write a comprehensive report on the GRI platform on its sustainability initiatives, sometimes exceeding a hundred pages, making it difficult for a detailed analysis of the report against other organizations.

To address the challenges associated with extracting information from these reports, we used the R programming language and a set of libraries to convert image-based reports (i.e., PDFs) into text documents. An optical character recognition (OCR) algorithm that recognizes the symbolic representation of texts, words, and layout information to extract a table of firm performance indicators is employed to extract a firm's sustainability practice. We saved detailed text descriptions of the sustainability practice indicators as separate documents for text mining. We coded whether a firm chose to report each of the six practices as a binary variable.

We then applied the latent Dirichlet allocation (LDA) based topic modeling algorithm to identify and summarize common themes expressed in a collection of documents [39]. A topic modeling is a machine algorithm designed to discover hidden thematic structures embedded in documents to organize and summarize a collection of documents. The model assumes that the distribution of topics and allocation of topics to documents follow a Dirichlet distribution. We identified three themes from the topic modeling analysis: Environment, social, and governance (ESG), and assigned probability scores for each document.

The data for this study comes from several sources. The first source is the annual sustainability reports of 969 firms across industries from 2013 to 2015 from the global reporting initiative (GRI). In addition, we matched this dataset with firms in the Compustat database to obtain variables about firm attributes such as firm size and firm profitability [16]. As a result, matched data on 680 firms are available in our merged dataset, and this is a strongly balanced panel dataset.

3.2. Models

3.2.1. Topic Modeling and Latent Markov Model

Using latent Dirichlet allocation (LDA) based topic modeling techniques, we identify three categories of strategic intent—(1) environment performance, (2) social engagement, and (3) governance—and assign the extent to which firms express such intent in the reports. In other words, our specified model includes three topic variables: Environmental

Sustainability **2021**, 13, 1069 7 of 15

discussion (E), social discussion (S), and governance discussion (G). The list of words that make up each topic is included in Table 1. We then model a firm's sustainability maturity as a function of its strategic intent using a latent Markov model (LMM) to estimate the statistical significance and the extent of their relationships.

Table 1. Topic modeling.

Topic 1: Governance	Topic 2: Social	Topic 3: Environment
business	report	product
custom	employ	energy
manage	manage	use
company	company	environment
develop	safety	water
value	train	emissive
global	group	operate
corporate	sustain	project
finance	work	total
sustain	risk	waste
service	supplier	plant
group	health	system
people	perform	reduce
response	inform	year
social	intern	power
provide	relate	material
innovation	response	increase
market	stakeholder	million
product	board	impact
active	number	improv

Our model investigates the maturity level of a firm's reporting sustainability practices. The model consists of six indicators as our dependent variables: (1) Economic (EC), (2) labor (LA), (3) product responsibility (PR), (4) environment (EN), (5) human rights (HR), and (6) social (SO), with each indicator as a categorical response of report (1) or no report (0). We use these six GRI indicators as a collective measure of a firm's sustainability maturity level. To better understand the factors that influence a firm's decision to report these measures, we identify themes expressed in a given year's report as covariates in the model. We further model a firm's intention behind its reporting decision as unobservable states, using an LMM approach. We use the R programming library LMest to estimate the optimal number of hidden states that lead to our model's best fit. We start with two hidden states to thirty states to accommodate possible reporting indicators up to six indicators. For instance, a firm's intention to report sustainability initiatives may be driven by the desire to attract and retain employees, or the goal of maintaining customer relations may drive it. Moreover, these intentions may switch over time. We further ran a model with three hidden states and three topics as covariates to estimate the three topics influencing the hidden states' statistical significance in initial and transition probabilities. We do not make any a priori assumptions about the number of intentions. We included the industry code as dummy variables for control variables.

3.2.2. Random Effects Models

Based on the results of topic modeling and LMM analysis, we observe many hidden states whose strategic intent is to manage external stakeholders. The customer relations management (CRM) hidden state reflects a firm's intent to manage firm reputation with customers, and based on its initial probability (0.21), external stakeholder management is most frequent within our sample. We also observe hidden states that reflect internal stakeholder management strategic intent. For instance, "workforce engagement" (WE) represents a firm's intent to attract and retain talent by disclosing firm activities associated with labor practices and human rights involvement. Last, the "environment performance"

Sustainability **2021**, 13, 1069 8 of 15

(ENV) hidden state is part of the workforce management and part of investors' management strategic intent.

Therefore, we further explore the influences of three topics discussed in a firm's annual sustainability report on the three types of firm's strategic intent and the path from these three strategic intents to the firm's financial performance. Variables used in the estimation models are listed in Table 2, and Table 3 shows the descriptive statistics of these variables.

Table 2. Description of key variables.

Variable	Description and Operationalization
Торіс Е	The variable is measured using data from GRI reports as the extent to which a firm describes environmental-related contents in its sustainability reports.
Topic S	The variable is measured using data from GRI reports as the extent to which a firm describes social-related content in its sustainability reports.
Topic G	The variable is measured using data from GRI reports as the extent to which a firm describes governance-related contents in its sustainability reports.
Reported Environment Motivation (REM)	Coded as 1 if a firm reports practices about sustainable resource use, pollution prevention, environment protection, and climate change mitigation and adaptation; otherwise coded as 0.
Reported Work Context (RWC)	Coded as 1 if a firm reports practices about employment, labor/management relations, training and education, diversity and equal opportunity, equal remuneration for women and men, supplier assessment for labor practices, and labor practice grievance mechanisms; otherwise coded as 0.
Reported Customer Orientations (RCO)	Coded as 1 if a firm reports practices about customer health and safety, product and service labeling, marketing communications, customer privacy, and compliance; otherwise coded as 0.
Firm Performance	Firm profitability. Log of ROA. ROA: Net income/total assets.
Firm Size Industry	Log of revenue. Industry category of a firm.

Table 3. Descriptive statistics.

	Observations	Mean	SD	Minimum	Maximum
Topic E	2016	0.18	0.18	0	0.87
Topic S	2016	0.21	0.20	0	0.78
Topic G	2016	0.19	0.18	0	0.89
REM	2016	0.26	0.44	0	1
RWC	2016	0.27	0.45	0	1
RCO	2016	0.29	0.46	0	1
Performance	1523	-0.18	1.66	-6.85	4.67
Size	1764	7.91	3.27	-4.42	16.90

First, for the estimation of the relationship between three topics discussed in a firm's annual sustainability report and three types of firm's strategic intent on the environment, work, and customer, since the dependent variables are binary measures, we run the random effects probit model:

$$y_{it} = \beta \mathbf{X}_{it} + e_{it}$$

where y is the probability of a firm's strategic intent on the environment, work, and customer, \mathbf{X} is a vector of independent variables, i and t are indices for individuals (firms) and time (year). β is a vector of parameters, and e_{it} are disturbances.

Sustainability **2021**, 13, 1069 9 of 15

We estimate the model that links strategic intent to firm performance, which is evaluated by $ln(ROA_{it})$, the natural log of the return on assets of firm i in year t. Therefore, we use a random-effects linear model for panel data, specified as:

$$y_{it} = \alpha_i + \mathbf{X}'_{it}\beta + \varepsilon_{it}$$

where y is the dependent variable, \mathbf{X} is a vector of independent variables, i and t are indices for individuals (firms) and time (year). β is a vector of parameters, α_i is the individual-specific effect that varies over i, and ε_{it} is the error term.

4. Results

As stated earlier, the methodology for this study followed a three-stage process. Accordingly, we have three sets of results. We first discuss the overall results and then delve deeper into the findings to provide further insights.

The first stage, involving the topic modeling, provided information about three categories of strategies intents: (1) Environment performance, (2) social engagement, and (3) governance, and assign the extent to which firms express such intent in the reports. In other words, our specified model includes three topic variables: Environmental discussion (E), social discussion (S), and governance discussion (G).

The second stage involved the latent Markov model (LMM) in providing initial probabilities associated with each hidden state. We derive from the model estimation as the strong motivation behind sustainability reporting and the three topics' distributions in each indicator document's detailed sections. The findings from this stage suggest that while most firms choose to disclose sustainability activities to manage relationships with external stakeholders, the internal workforce engagement strategic intent is gaining acceptance. In other words, firms disclose their sustainability reports to manage their reputation with customers. Firms also try to manage employee satisfaction through disclosure related to labor practices and human rights involvement. In addition, firms disclose environmentrelated activities to manage the customer relationship and investor relationship. We used the transition probabilities as selected results with topics discussed in a firm's annual sustainability report as the main variables of interest in influencing a firm's strategy switching behavior. These results suggest that based on the transition probabilities, more discussion on social engagement activities is likely to increase a firm's strategy switching behavior to other hidden states, showing an increase in its maturity level. Furthermore, the discussion of both environment and governance issues is likely to decrease WCs switching behavior, indicating that firms are likely to remain within the ENV strategy.

In the third stage, we found that the linear model estimation suggests that sustainability initiatives on environmental issues can increase firms' financial performance. Compared to the external intent, such as customer-oriented activities, internal intent, such as employee-oriented activities, significantly influences firm performance. Furthermore, the non-linear probit estimation suggested that all three topics have strong and positive influences on firms' strategic intents on the environment, work, and customer. We discuss these results further to provide granular insights.

4.1. Initial and Transition Probabilities

Table 4 shows the initial probabilities of each hidden state, which we derive from the model estimation as the strong motivation behind sustainability reporting and the three topics' distributions in each indicator document's detailed sections. First, we observe the "no disclosure" (ND) hidden state in which firms are unlikely to report any indicators or are simply terminating reporting practices. Unfortunately, most companies are likely to start in the no disclosure state (0.36) in our sample. We observe several hidden states whose strategic intent is managing external stakeholders. The customer relations management (CRM) hidden state reflects a firm's intent to manage firm reputation with customers, and based on its initial probability (0.21), external stakeholder management is most frequent within our sample. We also observe hidden states that reflect internal stakeholder man-

Sustainability **2021**, 13, 1069 10 of 15

agement strategic intent. For instance, "workforce engagement" (WE) represents a firm's intent to attract and retain talent by disclosing firm activities associated with labor practices and human rights involvement. Last, the "environment performance" (ENV) hidden state is part of the workforce management and investor management strategic intent.

Table 4. Initial probability.

Hidden State Labels	Initial Probabilities	Topic (E)	Topic (S)	Topic (G)
No disclosure (ND)	0.36	0.00	0.00	0.00
Customer relations management (CRM)	0.21	0.29	0.36	0.30
Environmental performance (ENV)	0.16	0.29	0.31	0.37
Workforce engagement (WE)	0.09	0.32	0.41	0.23

Tables 5 and 6 show selected results with topics discussed in a firm's annual sustainability report as the main variables of interest in influencing a firm's strategy switching behavior. Table 5 shows transition probabilities for ND as a prior hidden state, and results are displayed as multinomial logit, with CRM as the base or reference. The coefficients are interpreted as the odds of choosing one of five hidden states relative to the base hidden state. For instance, Table 5, showing a positive coefficient of 3.19 in the WC column, indicates that firms whose prior hidden state was ND are more likely to switch to WC, relative to CRM. Positive and statistically significant coefficients in social discussion (Topic S) rows indicate that more discussion on social engagement activities are likely to increase firm's strategy switching behavior to WC, relative to CRM.

Table 5. Transition probability for no disclosure (ND).

Base = CRM	WC	ENV
Intercept	-8.52 ***	-4.07 ***
Topic E	-1.18	-1.24
Topic S	3.19 ***	3.26 ***
Topic G	-0.80	-0.97

^{*} *p* < 0.1; ** *p* < 0.05; *** *p* < 0.01.

Table 6. Transition probability for environmental performance (ENV).

Base = ENV	WC	CRM
intercept	0.22	0.24
Topic E	-1.31 ***	-1.08
Topic S	-4.42 ***	1.07
Topic G	-2.34 ***	1.41
•	Industry Controls Included	

^{*} *p* < 0.1; ** *p* < 0.05; *** *p* < 0.01.

Table 6 further improves the model by including industry code as dummies and shows selected results for estimating the effects of three topics on switching behaviors of firms whose prior hidden state is ENV. The results show that the discussion of both environment and governance issues is likely to decrease WCs switching behavior, indicating that firms are likely to remain within the ENV strategy.

4.2. Econometric Estimation Results

We present the main analysis results of the probit model and linear model in Tables 7 and 8, respectively. From the results in Table 7, we find the strongly significant and positive re-

Sustainability **2021**, 13, 1069

lationship between three topics on environment, social, governance, and three strategic intents: REM, RWC, and RCO.

Table 7. Main estimation results for the probit model.

	(1)	(3)	
_	REM	RWC	RCO
Tomia E	20.483 ***	27.450 ***	12.839 ***
Topic E	(1.453)	(1.617)	(1.223)
Tarria C	4.970 ***	10.141 ***	24.150 ***
Topic S	(1.382)	(1.420)	(1.330)
TC	11.211 ***	28.937 ***	23.709 ***
Topic G	(1.774)	(1.589)	(1.404)
Cino	-0.025	0.195 ***	0.279 ***
Size	(0.072)	(0.064)	(0.061)
II I 1	-1.242	-0.973	3.458 ***
Heavy Ind.	(0.915)	(0.676)	(0.825)
Tanistina	1.684	0.553	-0.882
Logistics	(1.280)	(1.172)	(1.136)
Commodities	-0.835	-3.288 ***	0.923
Commodities	(0.883)	(0.681)	(0.709)
Managhatania	0.698	0.164	0.307
Manufacturing	(0.601)	(0.440)	(0.463)
Leisure	-0.124	1.161	1.095
Leisure	(1.198)	(0.952)	(1.102)
Constant	-12.741 ***	-23.913 ***	-23.445 ***
Constant	(1.250)	(1.002)	(0.897)
Observations	1764	1764	1764
No. of Firms	647	647	647
Wald χ^{2-}	203.14 ***	740.06 ***	932.68 ***

Standard errors in parentheses: * p < 0.1; ** p < 0.05; *** p < 0.01.

Table 8. Main estimation results for the linear model.

	(1)	(2)	(3)
	Performance	Performance	Performance
REM	0.079 * (0.044)		
RWC		0.082 * (0.047)	
RCO			0.058 (0.050)
Size	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Heavy Ind.	0.358 (0.551)	0.355 (0.549)	0.354 (0.550)
Logistics	-1.372 (0.873)	-1.364 (0.871)	-1.363 (0.872)
Commodities	-0.164 (0.490)	-0.163 (0.491)	-0.167 (0.490)

Sustainability **2021**, 13, 1069

Table 8. Cont.

	(1)	(2)	(3)
	Performance	Performance	Performance
Manufacturing	-0.026	-0.026	-0.024
	(0.349)	(0.350)	(0.349)
Leisure	-1.048 * (0.623)	-1.050* (0.624)	-1.053 * (0.626)
Constant	7.759 ***	7.757 ***	7.763 ***
	(0.221)	(0.220)	(0.220)
Observations	1450	1450	1450
No. of Firms	539	539	539

Robust standard errors in parentheses: * p < 0.1; ** p < 0.05; *** p < 0.01.

Estimation results with heteroskedasticity-consistent robust standard errors are presented in Table 8. All models have the same dependent variable, i.e., firm performance, while the dependent variables are different types of sustainable strategy. Column 1 presents the results of the impact of REM on the firm performance, column 2 presents the results of the impact of RWC on the firm performance, and column 3 presents the results of the impact of RCO on the firm performance. We find that while REM (β = 0.079, p < 0.1) and RWC (β = 0.082, p < 0.1) have a positive and significant relationship with firm performance, such influence of RCO on performance is not significant.

5. Discussion

5.1. Revisiting the Aim of the Paper

Through voluntary disclosure of sustainability activities on online reporting platforms, firms want to showcase their sustainability-related strategies to stakeholders. This study aims to explore the strategic intents embedded in firms' sustainability reports and evaluates the consequent impacts on firm performance. With growing sustainability data on online reporting platforms, data analytics offers opportunities to study the unstructured sustainability data and better understand firms' sustainability strategies and activities. This study applies a mixed-method approach to generate business intelligence based on text analysis of firms' sustainability reports. We explore the topics discussed in the sustainability reports using text analysis and topic modeling. Through the latent Markov model, we examine the rationale underlying the voluntary disclosure by firms. We further test the relationship between three topics and three types of strategic intent and the impacts of strategic sustainability intents on firm performance.

5.2. Findings of the Study

We have three key findings. First, we find that while most firms choose to disclose sustainability activities to manage relationships with external stakeholders, internal workforce engagement strategic intent is gaining acceptance. In other words, firms disclose their sustainability reports to manage their reputation with customers. Firms also try to manage employee satisfaction through disclosure related to labor practices and human rights involvement. In addition, firms disclose environment-related activities to manage customer relationships and investor relationships.

Second, the three major topics embedded in firms' sustainability reporting influence firms' strategy switching behavior. Based on the transition probabilities, we find that more discussion on social engagement activities is likely to increase a firm's strategy switching behavior to other hidden states, showing an increase in its maturity level. Furthermore, the discussion of both environment and governance issues is likely to decrease WCs switching behavior, indicating that firms are likely to remain within the ENV strategy. In addition, all three topics have a strong and positive influence on firms' strategic intents on the environment, work, and customer.

Sustainability **2021**, 13, 1069 13 of 15

Third, sustainability initiatives on environmental issues can increase firms' financial performance. Compared to the external intent, such as customer-oriented activities, internal intent, such as employee-oriented activities, significantly influences firm performance.

5.3. Theoretical and Scholarly Contribution of the Study

As noted above, the paper's aims, methods, and findings are aligned to the exploration of voluntary disclosure of sustainability activities on online reporting platforms that would help firms want to showcase their sustainability-related strategies to stakeholders. In this context, our study in one direction extends, and in other aspects, differs from the existing literature in several ways that we discuss further.

With all these findings, this study offers several theoretical contributions. First, it contributes to expanding the research stream of sustainability and green IS areas using data, applications, and systems approach, which is more informative and insightful. Several researchers focused on the quality of sustainability reports. They used statistical models to analyze the factors that explain how a company assures their sustainability report [40] and machine learning approach to develop a corporate sustainability report scoring solution [41].

However, previous research mostly only focused on one industry or organization or lack of discussions of contents embedded in corporate sustainability reporting. Our research addresses the research gap by focusing on diverse global companies' sustainability report and using the text mining approach and regression analysis to gain insights from the online sustainability reporting.

Second, some of the prior studies applied text mining to identify sustainability trends of companies in chemistry industries [42], others conducted a content analysis of corporate sustainability reports [43] or sustainability reporting by one public organization [44]. In this context, specific industries have different dynamics regarding disclosure activities, i.e., sometimes disclosure helps, and sometimes it hurts to gain stakeholders' confidence. A plausible explanation for that is how information is disclosed, i.e., only specific issue-based disclosures such as responsible behaviors, or responses to specific environmental behaviors, such as oil spills and deforestations. On the contrary, as this study suggests, both holistic or top-level and "information-derived disclosures" may enhance stakeholders' impression about a firm. This study provides a method or application to follow such an approach and enables a method to gain an in-depth understanding of firms' disclosure activities. This study complements some prior research findings that environment consideration is not the only motivation for voluntary disclosure [12]. Firms also try to demonstrate a sustainability path through customer and employee relationship management.

Third, this study also suggests that the strategic intents embedded in firms' sustainability reports are not static. We use transition probabilities to present these dynamics, providing a new research avenue. As prior research indicates, the sustainability reports reflect companies' sustainability emphases and interests and the companies' overall sustainability condition [45]. Aligned and extending the existing work, this study has consistent viewpoints and findings, providing more in-depth and more comprehensive understandings of the strategic intents embedded in firms' sustainability reports and the impacts on firm performance.

5.4. Conclusions

In conclusion, the key findings suggest that it is useful to investigate firms' sustainability strategic intent based on the critical topics they disclosed in their sustainability reports. In providing these insights, our study helps move the existing research towards a better understanding of business sustainability issues through analytics. Different stakeholders on the online reporting platform can get some practical implications from this study. For firms who disclose their sustainability reports, this study suggests that the disclosure of environmental sustainability initiatives and customer-oriented sustainability activities may bring value to their financial performance. For firms who read these sus-

Sustainability **2021**, 13, 1069 14 of 15

tainability reports to make investment decisions, they may focus more on the discussions around social engagement activities as it is a signal of a higher maturity level of firms' sustainability strategy.

Author Contributions: Conceptualization, D.Y. and X.N.; methodology, D.Y. and X.N.; software, D.Y. and X.N.; validation, J.K., X.N., and D.Y.; formal analysis, X.N. and D.Y.; investigation, J.K., and X.N.; resources, D.Y., X.N., and J.K.; data curation, D.Y. and X.N.; writing—original draft preparation, D.Y.; writing—review and editing, J.K. and X.N.; visualization, X.N. and D.Y.; supervision, J.K.; project administration, J.K.; funding acquisition, N/A; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.globalreporting.org/.

Acknowledgments: Authors acknowledge the help of Xueying Liu in data collection and preparation at the early stage of this project.

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

References

- 1. Arikan, E.; Kantur, D.; Maden, C.; Telci, E.E. Investigating the mediating role of corporate reputation on the relationship between corporate social responsibility and multiple stakeholder outcomes. *Qual. Quant.* **2016**, *50*, 129–149. [CrossRef]
- 2. Ocean Tomo. Ocean Tomo Announces Result of Annual Study of Intangible Asset Market Value. 2020. Available online: https://www.oceantomo.com/intangible-asset-market-value-study/ (accessed on 23 November 2020).
- 3. Avetisyan, E.; Hockerts, K. The consolidation of the ESG rating industry as an enactment of institutional retrogression. *Bus. Strategy Environ.* **2017**, *26*, 316–330. [CrossRef]
- 4. Gholami, R.; Watson, R.T.; Hasan, H.; Molla, A.; Bjorn-Andersen, N. Information systems solutions for environmental sustainability: How can we do more? *J. Assoc. Inf. Syst.* **2016**, *17*, 521–536. [CrossRef]
- 5. Malhotra, A.; Melville, N.P.; Watson, R.T. Spurring impactful research on information systems for environmental sustainability. *MIS Q.* **2013**, *37*, 1265–1274. [CrossRef]
- 6. Dhanorkar, S. Environmental benefits of internet-enabled C2C closed-loop supply chains: A quasi-experimental study of Craigslist. *Manag. Sci.* **2019**, *65*, *660–680*. [CrossRef]
- 7. Jacobs, B.W.; Singhal, V.R.; Subramanian, R. An empirical investigation of environmental performance and the market value of the firm. *J. Oper. Manag.* **2010**, *28*, 430–441. [CrossRef]
- 8. Dao, V.; Langella, I.; Carbo, J. From green to sustainability: Information Technology and an integrated sustainability framework. *J. Strateg. Inf. Syst.* **2011**, *20*, 63–79. [CrossRef]
- 9. Scholtens, B. Indicators of responsible investing. Ecol. Indic. 2014, 36, 382–385. [CrossRef]
- 10. Dyllick, T. Responsible management education for a sustainable world: The challenges for business schools. *J. Manag. Dev.* **2015**, 34, 16–33. [CrossRef]
- 11. Khan, N.; Sikes, J. IT Under Pressure: McKinsey Global Survey Results; McKinsey & Company: Hong Kong, China, 2014; pp. 1-8.
- 12. Taylor, J.; Vithayathil, J.; Yim, D. Are corporate social responsibility (CSR) initiatives such as sustainable development and environmental policies value enhancing or window dressing? *Corp. Soc. Responsib. Environ. Manag.* **2018**, 25, 971–980. [CrossRef]
- 13. Orlitzky, M. Corporate social responsibility, noise, and stock market volatility. Acad. Manag. Perspect. 2013, 27, 238–254. [CrossRef]
- 14. Eccles, R.G.; Ioannou, I.; Serafeim, G. The impact of corporate sustainability on organizational processes and performance. *Manag. Sci.* **2014**, *60*, 2835–2857. [CrossRef]
- 15. McWilliams, A.; Siegel, D. Corporate social responsibility and financial performance: Correlation or misspecification? *Strateg. Manag. J.* **2000**, *21*, 603–609. [CrossRef]
- 16. Nishant, R.; Teo, T.S.; Goh, M. Do shareholders value green information technology announcements? *J. Assoc. Inf. Syst.* **2017**, *18*, 542–576. [CrossRef]
- 17. Margolis, J.D.; Walsh, J.P. Misery loves companies: Rethinking social initiatives by business. *Adm. Sci. Q.* **2003**, *48*, 268–305. [CrossRef]
- 18. Robinson, S.R.; Irmak, C.; Jayachandran, S. Choice of cause in cause-related marketing. J. Mark. 2012, 76, 126–139. [CrossRef]
- 19. Sheikh, S.-U.-R.; Beise-Zee, R. Corporate social responsibility or cause-related marketing? The role of cause specificity of CSR. *J. Consum. Mark.* **2011**, *28*, 27–39. [CrossRef]
- 20. Van den Brink, D.; Odekerken-Schröder, G.; Pauwels, P. The effect of strategic and tactical cause-related marketing on consumers' brand loyalty. *J. Consum. Mark.* **2006**, 23, 15–25. [CrossRef]
- 21. Bhattacharya, C.B.; Sen, S.; Korschun, D. Using corporate social responsibility to win the war for talent. *MIT Sloan Manag. Rev.* **2008**, 49, 37–44.

Sustainability **2021**, 13, 1069 15 of 15

22. Kowal, J.; Roztocki, N. Job satisfaction of IT professionals in Poland: Does business competence matter? *J. Bus. Econ. Manag.* **2015**, 16, 995–1012. [CrossRef]

- 23. Edmans, A. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *J. Financ. Econ.* **2011**, 101, 621–640. [CrossRef]
- 24. Butler, T. Compliance with Institutional Imperatives on Environmental Sustainability: Building Theory on the Role of Green IS. *J. Strateg. Inf. Syst.* **2011**, *20*, 6–26. [CrossRef]
- 25. Chan, R.Y.; Ma, K.H. Impact of Executive Compensation on the Execution of It-Based Environmental Strategies under Competition. *Eur. J. Inf. Syst.* **2017**, *26*, 489–508. [CrossRef]
- 26. Wunderlich, P.; Veit, D.J.; Sarker, S. Adoption of Sustainable Technologies: A Mixed-Methods Study of German Households. *MIS Q.* **2019**, 43, 673–691. [CrossRef]
- 27. Melville, N.P. Information systems innovation for environmental sustainability. MIS Q. 2010, 34, 1–21. [CrossRef]
- 28. Bengtsson, F.; Ågerfalk, P.J. Information Technology as a Change Actant in Sustainability Innovation: Insights from Uppsala. *J. Strateg. Inf. Syst.* **2011**, 20, 96–112. [CrossRef]
- 29. Bose, R.; Luo, X. Integrative Framework for Assessing Firms' Potential to Undertake Green It Initiatives Via Virtualization—A Theoretical Perspective. *J. Strateg. Inf. Syst.* **2011**, *20*, 38–54. [CrossRef]
- 30. Loock, C.-M.; Staake, T.; Thiesse, F. Motivating Energy-Efficient Behavior with Green Is: An Investigation of Goal Setting and the Role of Defaults. *MIS Q.* **2013**, *37*, 1313–1332. [CrossRef]
- 31. Seidel, S.; Recker, J.; Vom Brocke, J. Sensemaking and Sustainable Practicing: Functional Affordances of Information Systems in Green Transformations. *MIS Q.* **2013**, *37*, 1275–1299. [CrossRef]
- 32. Cooper, V.; Molla, A. Information Systems Absorptive Capacity for Environmentally Driven Is-Enabled Transformation. *Inf. Syst. J.* 2017, 27, 379–425. [CrossRef]
- 33. Hasan, H.; Smith, S.; Finnegan, P. An Activity Theoretic Analysis of the Mediating Role of Information Systems in Tackling Climate Change Adaptation. *Inf. Syst. J.* **2017**, 27, 271–308. [CrossRef]
- 34. Zhu, Q.; Sarkis, J.; Lai, K.-H. Confirmation of a measurement model for green supply chain management practices implementation. *Int. J. Prod. Econ.* **2008**, *111*, 261–273. [CrossRef]
- 35. Corbett, C.J. How sustainable is big data? Prod. Oper. Manag. 2018, 27, 1685–1695. [CrossRef]
- 36. Ketter, W.; Peters, M.; Collins, J.; Gupta, A. A multiagent competitive gaming platform to address societal challenges. *MIS Q.* **2016**, *40*, 447–460. [CrossRef]
- 37. Ketter, W.; Peters, M.; Collins, J.; Gupta, A. Competitive benchmarking: An IS research approach to address wicked problems with big data and analytics. *MIS Q.* **2016**, *40*, 1057–1080. [CrossRef]
- 38. Nishant, R.; Kennedy, M.; Corbett, J. Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *Int. J. Inf. Manag.* **2020**, *53*, 1–13. [CrossRef]
- 39. Blei, D.M. Probabilistic topic models. Commun. ACM 2012, 55, 77–84. [CrossRef]
- 40. Fernandez-Feijoo, B.; Romero, S.; Ruiz, S. Multilevel Approach to Sustainability Report Assurance Decisions. *Aust. Account. Rev.* **2015**, 25, 346–358. [CrossRef]
- 41. Shahi, A.M.; Issac, B.; Modapothala, J.R. Intelligent Corporate Sustainability Report Scoring Solution Using Machine Learning Approach to Text Categorization. In Proceedings of the 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT), Kuala Lumpur, Malaysia, 6–9 October 2012; pp. 227–232.
- 42. Liew, W.; Adhitya, A.; Srinivasan, R. Sustainability Trends in the Process Industries: A Text Mining-Based Analysis. *Comput. Ind.* **2014**, *65*, 393–400. [CrossRef]
- 43. Landrum, N.E.; Ohsowski, B. Identifying Worldviews on Corporate Sustainability: A Content Analysis of Corporate Sustainability Reports. *Bus. Strategy Environ.* **2018**, 27, 128–151. [CrossRef]
- 44. An, Y.; Davey, H.; Harun, H. Sustainability Reporting at a New Zealand Public University: A Longitudinal Analysis. *Sustainability* **2017**, *9*, 1529. [CrossRef]
- 45. Zakhrani, K.K.; Kamayanti, A. Understanding of Sustainability Report: Semiotic Approach. J. Ilm. Mhs. FEB 2012, 1, 1–11.