15 REVISTA Kawsaypacha SOCIEDAD y MEDIO AMBIENTE

INTE PUCP

N° 15 enero – junio 2025. E-ISSN: 2709 – 3689

Revisión sistemática de literatura

Emerging Technologies for Socioenvironmental Auditing: Identification of Factors, Challenges and Technologies Using Text Mining and Analysis

Tecnologías emergentes para la auditoría socioambiental: identificación de factores, desafíos, tecnologías mediante minería de textos y análisis

Dr. Kamakshaiah Musunuru

Gandhi Institute of Technology and Management (GITAM Deemed to be University). School of Business, India

How to cite: Musunuru, K. (2025). Emerging Technologies for Socioenvironmental Auditing: Identification of Factors, Challenges and Technologies Using Text Mining and Analysis. Revista Kawsaypacha: Sociedad Y Medio Ambiente, (15), A-004. https://doi.org/10.18800/kawsaypacha.202501.A004

Abstract: Both the environment and society are fundamental aspects for businesses. Companies should consider these aspects while addressing challenges like transparency, accountability, sustainability and governance. This study aims at assessing the impact of information systems on Socioenvironmental Auditing Practices (SAP). For that purpose, it assumes that the impact of business activities on society and the environment can be audited, and the efficiency of such auditing practices depends on information systems. Datasets were mined from Scopus and Web of Science (WoS) using text mining methodology and were analyzed using the following tools: Correspondence Analysis (CA) to identify factors, challenges and technologies; Exploratory Factor Analysis (EFA) to identify underlying factor structures; Confirmatory Factor Analysis (CFA) to evaluate cause-and-effect relationships among factors, technologies and challenges. A few emerging technologies, viz., big data, blockchain, cloud computing solutions, machine learning, appear to have both a positive and a negative influence on SAP, but these are mostly insignificant. Challenges such as transparency, accountability, and sustainability were found to be associated with SAP. Blockchain technologies influence social and environmental factors positively, while machine learning influences EFA negatively yet its impact on CFA is positive and significant.

Keywords: Socioenvironmental auditing. Information systems. Big data technology. Text mining and analysis. Global challenges.

Resumen: El medioambiente y la sociedad son dos aspectos fundamentales para los negocios. Las empresas deben considerar estos aspectos al abordar desafíos como la transparencia, la responsabilidad, la sostenibilidad y la gobernanza. El objetivo de este estudio es evaluar el impacto de los sistemas de información sobre las prácticas de auditoría socioambiental (PAS) y parte de la premisa de que el impacto de las actividades de una organización en la sociedad y el medioambiente puede ser evaluado mediante prácticas de auditoría, cuya eficiencia depende de los sistemas de información. Los conjuntos de datos fueron extraídos de Scopus y Web of Science (WoS) utilizando una metodología de minería de textos y fueron analizados mediante Análisis de Correspondencia (AC) para identificar factores, desafíos y tecnologías; Análisis Factorial Exploratorio (AFE) para identificar estructuras subyacentes de factores; y Análisis Factorial Confirmatorio (AFC) para evaluar las relaciones de causa y efecto entre factores, tecnologías y desafíos. Algunas tecnologías emergentes, como big data, blockchain, soluciones en la nube y aprendizaje automático, parecen influir en las PAS tanto de manera positiva como negativa, pero en su mayoría de manera insignificante. Se encontró que los desafíos como la transparencia, la responsabilidad y la sostenibilidad están asociados con las PAS. La tecnología blockchain influye positivamente en los factores sociales y ambientales, mientras que el aprendizaje automático influye negativamente en el AFE, aunque su impacto es positivo y significativo en el AFC.

Palabras clave: Auditoría socioambiental. Sistemas de información. Tecnología de big data. Minería y análisis de texto. Desafíos globales.

1. Introduction

The main aim of this paper is to assess the impact of information systems on SAP, particularly that of big data and its associated technologies. The natural environment and society are two fundamental aspects to consider when doing business. Firms should be sensitive to these aspects, so that they can implement governance mechanisms to develop transparent, accountable, and sustainable practices (Wong et al., 2021). SAP are crucial for the assessment of business activities and their impact on both society and environment (Oberst & Gheorghita, 2022; Vinšalek-Stipić, 2022). These practices are key for identifying and addressing organizational challenges such as transparency, sustainability, and accountability (Montero & Le Blanc, 2019; Sysoieva et al., 2023; Shahib et al., 2020). This study evaluates the aforementioned challenges within the context of Socioenvironmental Auditing (SA), considering the mediation by emerging technologies.

Following current literature, SA is conducted through various methodologies, such as environmental impact assessments, social impact assessments, life cycle assessments, and sustainability reporting frameworks, such as the Global Reporting Initiative (GRI) guidelines (Demirel et al., 2020; Sucena et al., 2019). However, the assessment of technology's impact on sustainability remains underexplored in existing literature.

Information Systems and Technology (IST) facilitate the collection of socioenvironmental data from various sources, including sensors, databases, surveys, and third-party sources (Cao et al., 2021; Troshani et al., 2022). These data may include environmental performance metrics, social impact indicators, and stakeholder feedback. Advanced analytics -e.g., big data and machine learning- allow auditors to analyze large volumes of socioenvironmental data to identify trends, patterns, and correlations. The current study assumes that information systems may be useful in assessing risks arising from organizational activities that impact the environment and society. Particularly, it has been found that cloud-based platforms help auditors to collect data in the field, conduct real-time assessments, and collaborate with stakeholders remotely, while blockchains are deemed to offer integrity and immutability to audit trails (Ghobakhloo et al., 2023).

A few studies highlight the role of Industry 4.0 and other emerging technologies in shaping socioenvironmental auditing practices. For instance, Yusoff et al. (2023), refers to challenges and opportunities that Industry 4.0 -which integrates advanced technologies, e.g. IoT, AI/ML-poses for socioenvironmental auditing. Nwachukwu et al. (2021), argue that a deeper understanding of Industry 4.0 is necessary, as it might assist social auditors in analyzing data, identifying risks, and adapting audit procedures (Tavares & Azevedo, 2022). Understanding the impact of Industry 4.0 technologies on social dimensions of sustainable performance is crucial for organizations, as these technologies can directly and indirectly influence sustainable development goals, such as responsible consumption and production, through tools such as big data analytics and cloud computing solutions (Bai et al., 2023). The current study assumes that big data processing technologies have an impact on socioenvironmental auditing practices and aims to investigate these influences using systematic and scientifically grounded research methods, such as text mining and analysis.

1.1 Big Data for Social Auditing

One of the main objectives of this study is evaluating the influence of big data processing technologies on socioenvironmental auditing. These tools have become critical for social auditing, since they allow organizations and governments to assess the impact of policies, programs, and initiatives on society (Blazquez & Domenech, 2018). Their contribution has led to enhancing audit methods, expanding the audit scope, and injecting vitality into national audit work (Ma, 2023). The integration of big data processing technologies in auditing has significantly improved work efficiency and quality compared to traditional methods, making it a crucial component for future audit practices (Zhang, 2023).

By and large, the literature appears to highlight policies, programs, innovation in auditing as far as big data processing technologies are concerned. However, most of the discussion focuses on Industry 4.0 and emerging technologies instead of on these tools. Thus, the impact of big data processing technologies on social auditing remains vague and unclear for the research community. Despite this, literature shows evidence of the potential impact of information systems on social auditing, and this study tests this hypothesis using confirmatory factor analysis in one of the forthcoming sections.

1.2 Big Data for Environmental Auditing

Big data has emerged as a powerful tool for environmental auditing, offering new ways to monitor, assess, and manage environmental resources and impacts (Rahmadhani et al., 2023). Environmental auditing involves the systematic examination and evaluation of an organization's or government's environmental performance and compliance with regulatory requirements. It aims at identifying environmental risks, assessing impacts, and recommending strategies for improvement, to ensure sustainable resource management and minimize negative environmental consequences (Dovgal & Kuizheva, 2022).

Big data processing technologies play a crucial role in environmental auditing by providing innovative methods for data-driven auditing processes and sustainable development initiatives (Rahmadhani et al., 2023; Wang, 2022). However, challenges do exist, such as the need for standardized sustainability audit requirements and the dependency on the number and quality of data provided by economic units for effective auditing. Overall, the integration of big data processing technologies in environmental auditing not only enhances efficiency but also fosters advancements in sustainability practices and environmental protection. The current study builds on these insights by proposing a couple of hypotheses and testing them through confirmatory factor analysis.

2. Research Methods

The current study is both exploratory and causal in nature. The study aims to explore socioenvironmental challenges/issues, as well as to seek suitable technology-based solutions to address them. For this purpose, data sets were mined from existing literature using text mining methodology. Regarding documents, abstracts were collected from popular databases such as Scopus and Web of Science (WoS) using the query – «big data analytics for socioenvironmental auditing» in March 2024. It was possible to retrieve 62 documents: 14 documents from Scopus and 48 documents from Web of Science.

2.1 Text mining and analysis

Text mining and analysis is the process of deriving meaningful insights and patterns from unstructured textual data. It involves techniques and algorithms to extract, process, and analyze large volumes of text data to uncover valuable information (Musunuru, 2024a). All the abstracts, which were collected from Scopus and WoS, were arranged in a 62 X 1 data matrix, which was used as the text corpus for analysis. Data mining and analysis was carried out using a «tm» package of R programming language and R scripts are available in the article’s companion repository at https://github.com/Kamakshaiah/big-data-socio-environmental-context/tree/main/scripts

The text corpus is used to find conceptual patterns by extracting word vectors and Document Term Matrices (DTM). The resultant datasets are available at

https://github.com/Kamakshaiah/big-data-socio-environmental-context/tree/main/data

The investigation for factors, challenges and technology enablers was carried out by analyzing conceptual patterns using Correspondence Analysis (CA), Exploratory and Confirmatory Factor Analysis (EFA & CFA).

3. Data analysis

This section comprises three different parts, each dealing with a specific statistical technique. The first part covers the Correspondence Analysis (CA), which is used to identify conceptual patterns and their association with challenges and technology-based enablers. The second part addresses the realization of factor structures comprised by conceptual patterns (word/term vectors) as observed variables. The third part deals with CFA to evaluate the cause-and-effect relationships among study factors, challenges/issues and technology-based enablers.

3.1 Correspondence analysis

The variant of CA used for this study is the simple (symmetric) correspondence analysis, which is applied on rows and columns (Beh & Lombardo, 2014). The analysis was done using R programming language (R Core Team, 2023).

Table 1. Correspondence Analysis Results

Concept/Dimension a	Contribution b		Coordinates c		Cos2 d		Inertia e
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2	X
Academic	3.663	0.354	1.222	-0.299	0.653	0.039	0.028
Accountability	1.434	0.155	1.060	0.274	0.282	0.019	0.025
Accounting	8.420	0.172	0.931	-0.105	0.408	0.005	0.102
Adoption	3.987	0.138	1.301	-0.190	0.597	0.013	0.033
Analysis	0.006	0.036	-0.031	-0.060	0.001	0.002	0.047
Analytical	0.003	0.003	0.069	-0.054	0.000	0.000	0.034
Analyze	0.222	0.026	0.417	-0.112	0.046	0.003	0.024
Application	0.561	0.919	-0.410	0.413	0.077	0.078	0.036
Applications	0.325	0.362	-0.284	-0.236	0.025	0.018	0.063
Approach	0.938	11.572	-0.857	2.369	0.035	0.270	0.131
Artificial	0.006	0.042	0.088	0.183	0.001	0.003	0.038
Audit	0.946	0.926	-0.296	0.230	0.044	0.026	0.107
Auditing	0.039	0.285	-0.074	-0.157	0.003	0.015	0.060
Auditors	0.003	0.136	-0.030	0.169	0.000	0.004	0.101
Big	3.913	0.952	-0.632	-0.245	0.180	0.027	0.107
Blockchain	0.663	0.035	0.491	-0.089	0.057	0.002	0.058
Business	0.005	0.597	0.045	-0.388	0.000	0.023	0.079
Challenges	0.213	0.016	-0.425	-0.091	0.012	0.001	0.088
Cloud	7.183	6.608	-1.617	-1.220	0.264	0.150	0.134
Communication	0.745	0.341	-0.974	-0.518	0.081	0.023	0.045
Data	13.597	3.667	-0.765	-0.312	0.493	0.082	0.136
Databases	2.618	0.380	1.185	-0.355	0.459	0.041	0.028
Decision	0.315	11.571	-0.382	1.821	0.014	0.311	0.114
Design methodology approach	0.478	0.001	0.698	0.028	0.148	0.000	0.016
Disruptive	0.373	0.008	0.650	-0.075	0.078	0.001	0.024
Domain	0.090	0.041	0.320	0.169	0.010	0.003	0.046
Economic	0.307	0.000	-0.417	-0.007	0.014	0.000	0.105
Efficiency	0.041	0.007	0.144	-0.047	0.004	0.000	0.054
Emerging	0.285	0.568	-0.568	-0.631	0.025	0.031	0.057
Employees	0.502	0.013	0.855	-0.107	0.212	0.003	0.012
Environment	0.030	0.035	0.148	0.126	0.003	0.002	0.046
environmental	0.073	0.078	0.198	-0.161	0.006	0.004	0.061
Financial	1.626	0.151	0.868	-0.208	0.378	0.022	0.021
Governance	0.076	0.126	0.207	0.210	0.008	0.008	0.047
Implications	2.170	0.006	0.889	-0.036	0.449	0.001	0.024
Industry	6.929	0.270	1.177	-0.183	0.601	0.014	0.057
Information	0.000	0.016	-0.009	0.044	0.000	0.002	0.030
Intelligence	0.006	0.042	0.088	0.183	0.001	0.003	0.038

Internet	0.238	0.139	-0.493	-0.296	0.027	0.010	0.043
Knowledge	0.156	0.147	0.445	-0.341	0.062	0.036	0.012
Learning	0.443	25.592	-0.568	3.395	0.019	0.662	0.118
Limitations	0.134	0.008	0.352	-0.069	0.038	0.001	0.017
Limitations implications	0.681	0.001	0.996	-0.032	0.238	0.000	0.014
Literature	5.451	0.225	1.164	-0.186	0.646	0.016	0.042
Machine	0.075	7.468	-0.277	2.170	0.007	0.402	0.057
Management	0.286	1.204	0.317	0.512	0.012	0.032	0.116
methodology	0.372	0.023	0.540	0.107	0.053	0.002	0.035
National	0.119	0.039	-0.305	-0.138	0.017	0.004	0.034
Network	0.430	9.932	-0.698	2.638	0.023	0.333	0.091
Networks	0.472	0.000	-0.829	0.001	0.050	0.000	0.046
Organization	1.145	0.167	0.986	-0.297	0.292	0.026	0.019
organizations	0.281	0.069	0.640	-0.250	0.058	0.009	0.024
Originality value	0.478	0.001	0.698	0.028	0.148	0.000	0.016
Performance	1.049	0.621	-0.793	0.480	0.137	0.050	0.038
Regulation	3.104	0.310	1.258	-0.312	0.541	0.033	0.028
Regulatory	0.428	0.030	0.790	-0.164	0.123	0.005	0.017
Research	0.347	0.226	0.235	-0.149	0.037	0.015	0.046
Risk	0.054	3.539	-0.155	0.985	0.003	0.121	0.089
Security	4.759	1.040	-1.088	-0.400	0.235	0.032	0.100
Services	1.385	0.128	-1.004	-0.240	0.097	0.006	0.070
Social	0.236	0.007	0.204	0.027	0.033	0.001	0.035
Society	0.769	0.014	0.723	0.076	0.165	0.002	0.023
Software	0.019	1.022	-0.167	0.960	0.002	0.067	0.046
Storage	3.342	2.912	-1.560	-1.145	0.261	0.141	0.063
sustainability	0.766	0.069	0.659	-0.156	0.123	0.007	0.031
System	0.886	0.266	-0.549	0.236	0.047	0.009	0.093
Systematic	1.842	0.006	1.083	0.050	0.403	0.001	0.023
Systems	0.095	1.033	0.162	0.419	0.013	0.084	0.037
Techniques	0.079	2.372	-0.271	1.166	0.006	0.103	0.070
Technologies	0.281	0.040	0.332	-0.099	0.031	0.003	0.044
Technology	0.202	0.027	-0.178	-0.051	0.016	0.001	0.064
Theoretical	0.488	0.000	-0.743	-0.001	0.055	0.000	0.043
Transparency	1.016	0.342	0.892	0.407	0.221	0.046	0.023
Trends	1.892	0.297	1.035	-0.322	0.486	0.047	0.019
Users	3.408	0.029	-1.389	-0.101	0.180	0.001	0.093

Source – Derived from correspondence analysis of text mining data.

Table 1 presents the results of the correspondence analysis for columns -i.e., term vectors- which are deemed as conceptual patterns in this study. The table depicts the concepts considering their contribution, coordinates, cos2 values and inertia. Concepts such as Big, Data, Cloud, Storage, Security, Regulation, Accountability, and Transparency contribute highly to Dimension 1. Similarly, Dimension 2 is highly influenced by concepts such as Decision, Machine, Learning, Cloud, Storage, Data, Risk, Network, Security, Systems and Transparency. Dimension 1 appears to be associated with technologies such as cloud storage and big data, while simultaneously presenting regulatory issues, accountability, and transparency as challenges. In contrast, Dimension 2 is aligned with technologies such as machine learning alongside cloud storage, and it presents security and transparency as challenges in big data governance and auditing. The inertia values, which indicate how much information is contained in each dimension based on the contributing concepts, suggest that Dimension 2 accounts for more variance. As a result, vectors such as machine learning, cloud storage, data storage, risk, security and transparency emerge as key theoretical concepts within the dataset. Table 2 outlines a group of potential factors, technologies and challenges identified through the correspondence analysis.

Table 2. Factors, challenges/issues and technology-based enablers

Quarter	Factors	Technologies	Challenges/issues
1	Environmental	Cloud storage	Big data
2	Environmental, Financial	Blockchain	Sustainability, regulatory compliance
3	Social, environmental	Digital technologies, information systems	transparency, accountability, methodology
4	Auditing	Information systems, machine learning	Governance, network, management

Source – Derived from correspondence analysis of text mining data.

Information systems and digital technologies seem to address challenges such as transparency, accountability and governance, which are closely linked to broader socioenvironmental concerns. However, certain technologies -such as cloud storage and blockchain- seem to address sustainability, enforcement and auditing issues, showing an environmental focus rather than a social one. Interestingly, machine learning emerges as a technology that is relevant to challenges related to auditing processes.

3.2 Exploratory Factor analys

Exploratory factor analysis (EFA) is a statistical method used to identify covariance structures in the datasets under analysis. It is a factor analysis technique whose overarching goal is to identify the underlying relationships between measured variables (Norris & Lecavalier, 2009). In this study, EFA was conducted using a *psych* package of R programming language (Revelle, 2020). This analysis was used to assess the variables priorly identified through correspondence analysis to uncover the underlying covariance structure.

The factors, technologies and challenges analyzed were essentially conceptual patterns, which -regardless of their characteristic value- were treated as study variables for further analysis (Musunuru, 2024a, 2024b). Moreover, the data collected for this study is representative of the research focus. Figure 1 presents the scree plot that resulted from EFA, and based on visual inspection, a two-factor solution appears to be the most viable choice for further analysis. Table 3 shows the factor loadings for term vectors, previously shown in Table 1.

Figure 1. Scree plot

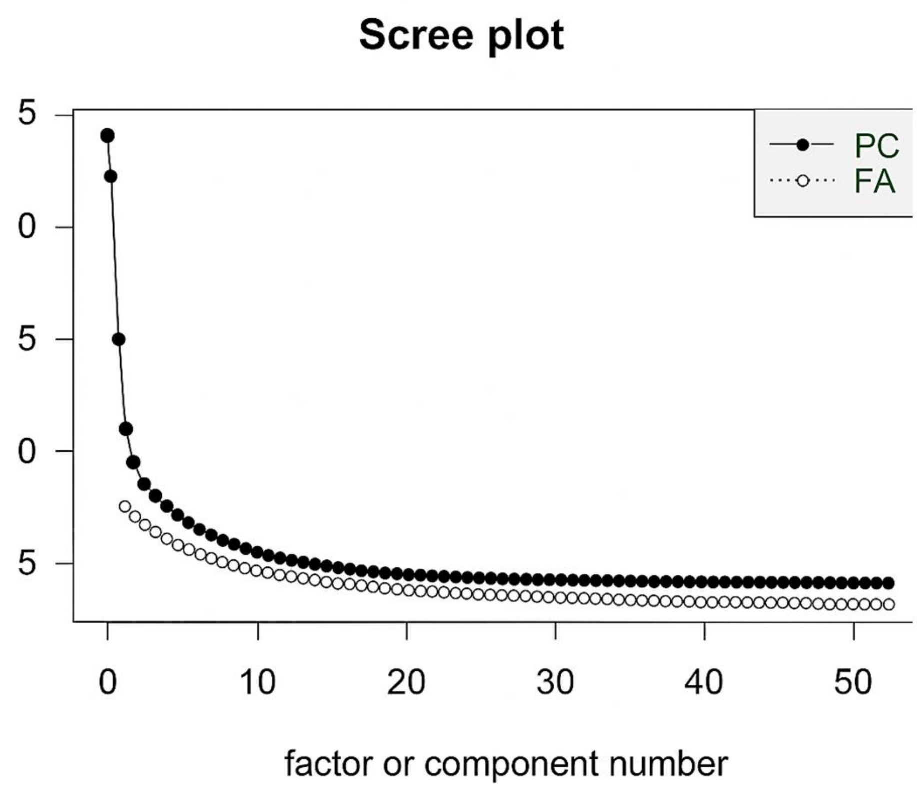


Table 3. Factor loadings

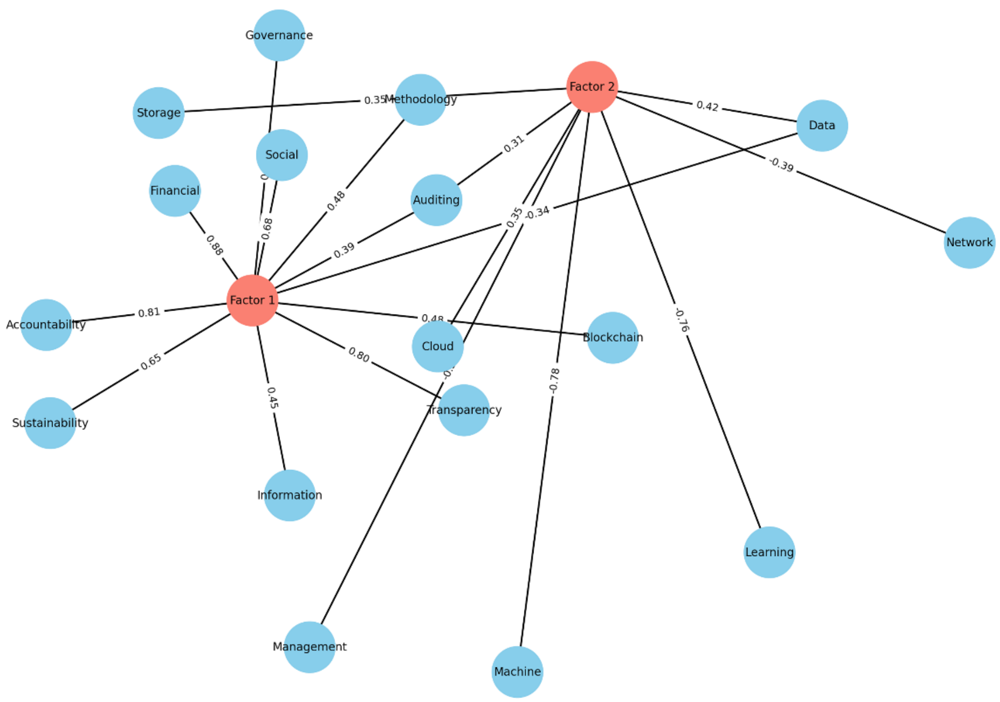
Variable	Factor 1	Factor 2
Environmental	0.297	0.144
Cloud	-0.280	0.351
Storage	-0.292	0.348
Big	-0.168	0.294
Data	-0.343	0.416
Financial	0.879	0.101
blockchain	0.480	0.072
sustainability	0.647	0.098
Social	0.678	-0.079
technology	0.060	0.251
information	0.451	0.245
Systems	-0.094	0.000
transparency	0.802	-0.104
accountability	0.813	-0.061
methodology	0.477	0.037
Auditing	0.385	0.309
Machine	-0.021	-0.784
Learning	-0.060	-0.763
Governance	0.395	0.009
Network	-0.089	-0.391
Management	0.100	-0.300

Source – Analysis of text mining data

Table 3 provides a summary of the factor analysis conducted on the conceptual patterns identified in this study. A close examination reveals there are two distinct dimensions that correspond to recurring ideas in literature. Factor 1 explains socioenvironmental challenges/issues. Several conceptual patterns are closely associated with this factor, e.g., financial (0.879), transparency (0.802), sustainability (0.647), governance (0.395) and accountability (0.813). Interestingly, patterns such as social (0.678), environmental (0.297), auditing (0.385) also load positively on this factor, together with technology-based enablers such as blockchain (0.480), information (0.451) and technology (0.060). However, other enablers such as big (-0.168), data (-0.343), systems (-0.094), machine (-0.021), learning (-0.060), cloud (-0.280) and storage (-0.292) load negatively on Factor 1. Though the literature supports the basic assumptions of this study, pointing out that blockchain is a technology that influences SA, it surprisingly fails to consider that technologies such as big data, cloud storage, and machine learning have an impact on it. Factor 2 appears to be more aligned with environmental challenges, rather than socioenvironmental ones. Interestingly, the technology enablers that fail to address socioenvironmental challenges/issues -namely, big (0.294), data (0.416), cloud (0.351), and storage (0.348)- seem to be relevant to address environmental challenges/issues.

The analysis results suggest that transparency and accountability are more strongly associated with social concerns, while governance covers both social and environmental challenges. Blockchain technology appears to positively influence both factors, while machine learning seems to have a negative influence on them. By and large, information systems appear to serve as enablers for both social and environmental concerns, while technologies such as big data and cloud storage act as solutions for environmental ones. Figure 2 provides a visual representation of this interpretation in relation to the structural model of the study.

Figure 2. Structural model



Source – Derived from the EFA using R language.

3.3 Confirmatory factor analysis

The current study assumes that SA can be influenced by big data processing technologies, and that such influence may be mediated by specific technologies. Literature supports the role of information systems as potential enablers of SA, particularly that of specific technologies such as machine learning, blockchain, cloud storage, etc. CFA was used to explain cause-and-effect relationships among study variables, and three candidate models were proposed:

1. The relationship between SA and big data processing technologies is mediated by information systems (H1).

2. The relationship between SA and big data processing technologies is mediated by machine learning (H2).

3. The relationship between SA and big data processing technologies is mediated by blockchain technology (H3).

Table 4. Summary of CFA

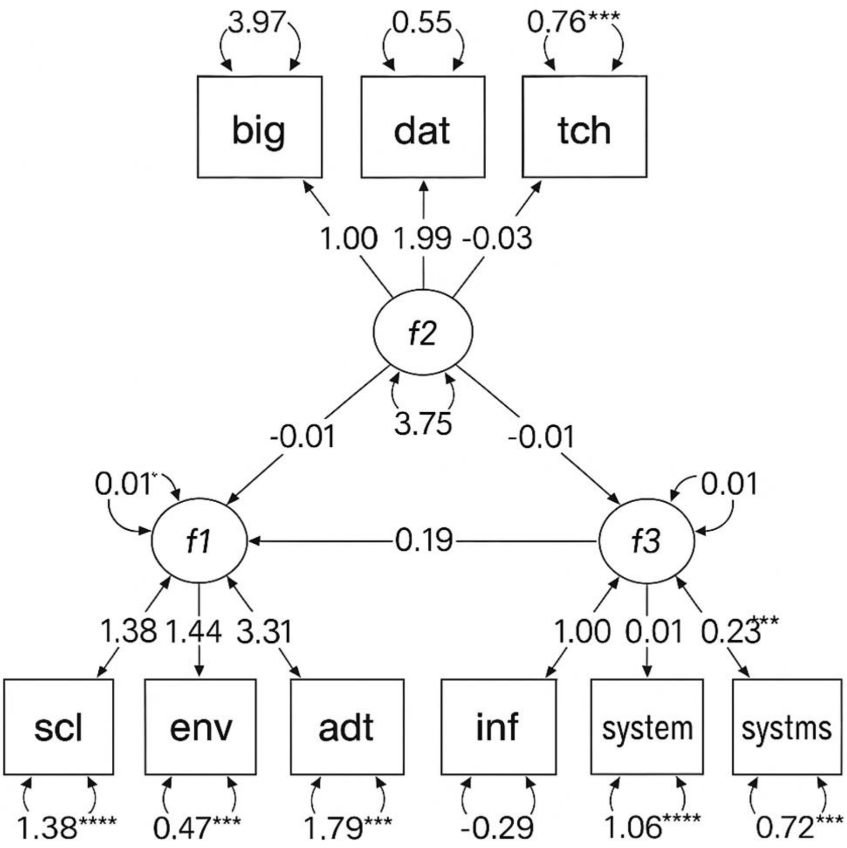
	lhs	op	rhs	Label	est	se	z	Pvalue	Constructs	Effect	Inference
Model 1	f1	~	f2	C	-0.037	0.028	-1.302	0.193	SA(f1)	Negative	Insignificant
	f3	~	f2	A	-0.005	0.072	-0.075	0.940	Big data processing technology (f2)	Negative	Insignificant
	f1	~	f3	B	0.191	0.144	1.326	0.185	Information systems - IS (f3)	Positive	Insignificant
	indirect	:=	a*b	Indirect	-0.001	0.014	-0.075	0.940	Mediation from IS	Negative	Insignificant
	total	:=	c+(a*b)	Total	-0.038	0.033	-1.137	0.255	The total effect	Negative	Insignificant
Model 2	f1	~	f2	C	-0.050	0.050	-1.006	0.315	SA(f1)	Negative	Insignificant
	f3	~	f2	A	-0.044	0.028	-1.560	0.119	Big data processing technology (f2)	Negative	Insignificant
	f1	~	f3	B	-0.313	0.304	-1.028	0.304	Machine Learning -ML (f3)	Negative	Insignificant
	indirect	:=	a*b	Indirect	0.014	0.016	0.877	0.380	Mediation from ML	Positive	Insignificant
	total	:=	c+(a*b)	Total	-0.037	0.040	-0.915	0.360	The total effect	Negative	Insignificant
Model 3	f1	~	f2	C	0.014	0.032	0.437	0.662	SA(f1)	Positive	Insignificant
	f3	~	f2	A	-0.005	0.013	-0.363	0.717	Big data processing technology (f2)	Negative	Insignificant
	f1	~	f3	B	1.689	0.444	3.805	0.000	Blockchain technology - BC (f3)	Positive	Significant
	indirect	:=	a*b	Indirect	-0.008	0.022	-0.364	0.716	Mediation from BC	Negative	Insignificant
	total	:=	c+(a*b)	Total	0.006	0.020	0.300	0.764	The total effect	Positive	Insignificant

Source – Derived from CFA of text mining data.

In Table 4, the columns labeled «lhs» and «rhs» refer to the left-hand side and right-hand side of the formula for a specific modeled relationship. The column «op» indicates the operator used in each of these relationships. In this column, the symbols «~» and «~~» represent regression and covariance, respectively. For instance, the formula «f1 ~ f2» refers to the relationship between «socioenvironmental auditing» and «Big data processing technologies». The symbol «: =» denotes the calculation made to determine direct and indirect effects associated with cause-and-effect relationships. Other columns, such as «label», «est», «se», «z» and «pvalue» refer to coefficients, estimates, standard error, z-statistic and p-value, respectively. In addition, the columns labeled as «Constructs», «effects» and «inference» provide information on the study constructs, the path estimates and concerned decisions, respectively. For instance, in Model 1, the overall estimate for cause-and-effect relationships is positive but insignificant. However, a closer look at specific relationships between pairs of variables reveals that big data processing technologies negatively influence both SA and information systems, although these effects are not statistically significant. Interestingly, information systems have a positive influence on SA, though once again such influence is not significant.

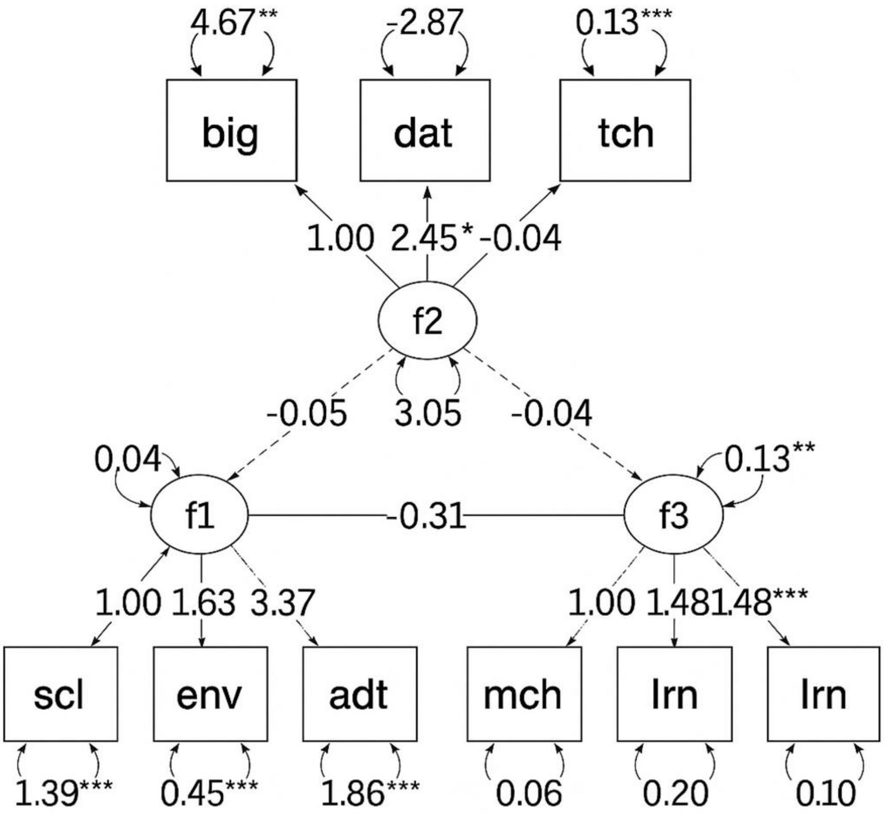
These findings clearly reveal that even though big data processing technologies may not appear in the literature as the main factor, information systems seem to play an influential role. The mediation effect of information systems is also negative but insignificant. Figure 3 provides visual representation of the meditation role played by Information Systems.

Figure 3. Meditation of Information Systems



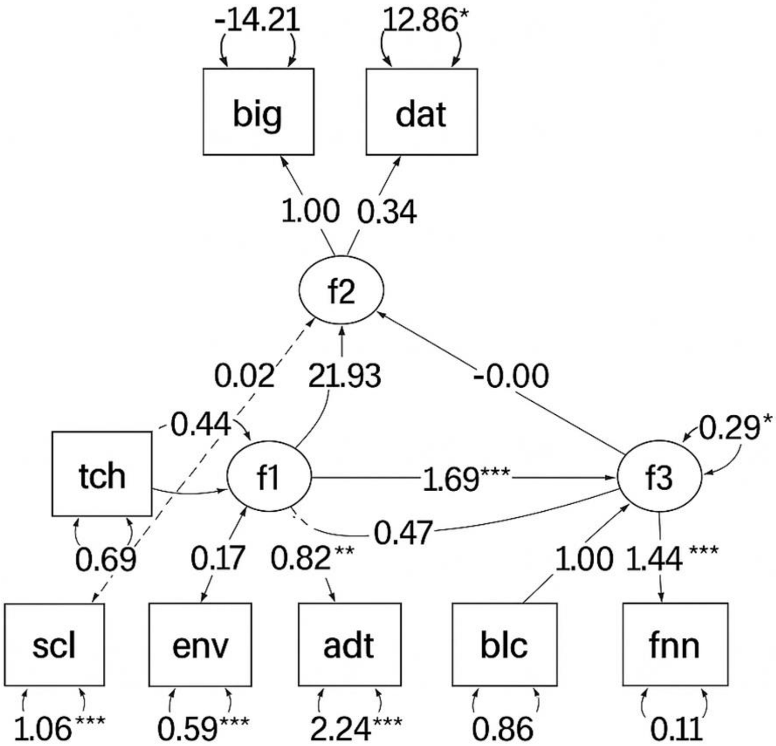
Model 2 examines the cause-and-effect relationships between SA and big data processing technologies, with machine learning acting as a mediating technology. The overall estimate for cause-and-effect relationships in Model 2 is negative but insignificant. Notwithstanding, a closer analysis of the relationships between pairs of variables reveals that big data processing technologies negatively influence both SA and machine learning, though these effects are not statistically significant. Interestingly, despite the fact that machine learning as a technology exerts a negative influence on SA, its mediation role appears to be positive. These results clearly suggest that though big data processing technologies do not appear to be the main factor in supporting SA, according to the existing literature, the research community certainly recognizes the role of machine learning as a relevant technology enabler for the other two constructs in this model. Figure 4 provides a visual representation of the meditation role played by Machine Learning.

Figure 4. Meditation of Machine Learning



Model 3 explores cause-and-effect relationships between SA and big data processing technologies, with fintech and blockchain technologies as mediators. The overall estimate for cause-and-effect relationships in Model 3 is positive but insignificant, while the mediating role played by fintech and blockchain technologies is negative and insignificant. However, when we examine the relationships between pairs of variables, big data processing technologies have a positive influence on SA, while fintech and blockchain have a negative yet insignificant influence on big data processing technologies. Remarkably, fintech and blockchain technologies influence SA positively, however their mediation role appears to be negative. These findings clearly indicate that, even though the current literature refers to big data processing technologies as the main factor supporting SA, the research community certainly agrees on the relevance of fintech and blockchain technologies as mediators between the other two constructs in the model. Model 3 is key, as it suggests that the SA – fintech/blockchain relationship is not only positive but statistically significant. Figure 5 provides visual representation of the meditation role played by blockchain technology.

Figure 5. Meditation of blockchain technology



4. Summary and discussion

The Correspondence Analysis suggests that information systems and digital technologies address challenges/issues such as transparency, accountability and governance, factors that are tightly connected to broader socioenvironmental concerns. This finding confirms what Jayasuriya et al., (2022), and Lamboglia et al (2021) pointed out, i.e., that social projects may be easily prepared and implemented with decentralized information systems. Furthermore, transparency and efficiency would contribute to changing organization culture, transforming employee interactions, both internally and with the wider community. Zhang (2023), also supports this conclusion in the study of information technologies and their application in resource and environmental auditing. Thus, these findings make clear that transparency and accountability are central challenges for both environmental and social dimensions. Moreover, the current study finds that managerial challenges are positively associated with auditing practices, a conclusion supported by Aswar et al. (2021), who observed that social factors have an impact on the administration and auditing practices, and may also discourage their efforts towards management reforms, regulatory compliance, and transparency culture. These insights clearly show that auditing practices are more closely connected to managerial challenges than to purely social and environmental ones.

The EFA further indicates that transparency and accountability are mainly social concerns, while governance is both a social and an environmental affair. Studies such as Jayasuriya et al., (2022) and Aswar et al. (2021), emphasize accountability as a social concern. Regarding technology enablers, blockchain technologies have a positive influence on both the social and environmental dimensions, while machine learning exerts a negative influence. Abreu et al. (2018) studied blockchain applications and their nexus with socioenvironmental auditing practices. Harber et al. (2023), also find that auditors attempt to affect policy outcomes to align with their firms' economic interests. As policies are both social and environmental in nature, it is plausible to establish that blockchain serves as a dual-purpose technology enabler for both dimensions. By and large, information systems are deemed as technology enablers for both social and environmental concerns, whereas big data and cloud storage seem to be more relevant for addressing environmental challenges, along with blockchain.

The CFA was conducted to test the statistical significance of these relationships. For this purpose, three candidate models were proposed and assessed. To a great extent, big data processing technologies appear to influence SAP negatively but insignificantly across all models. However, the mediation analysis provided more interesting results:

Model 1 – A partial mediation role of information systems (IS) was found, suggesting that IS positively influence SAP (This conclusion is consistent with Jayasuriya et al., 2022; Lamboglia et al 2021; Zhang et al., 2013). However, IS was found to negatively -though insignificantly- impact big data technologies, following SA literature. This inconsistency calls for further investigation and represents a potential avenue for future research.

Model 2 – The mediation role of machine learning was found to be positive but statistically insignificant. This conclusion is consistent with Benbouzid, B. (2019) and Li (2021), who refer to machine learning applications for socioenvironmental auditing. Even though the impact was positive, the evidence is insufficient. Thus, further investigation is required, and this study opens the door for future research.

Model 3 – While big data processing technologies do not appear to be influenced by blockchain, both technologies have a positive influence on SAP. The mediation role of blockchain is both positive and statistically significant, offering clear evidence and lucid conclusions on its relevance in socioenvironmental auditing.

5. Conclusion

This study used text mining, Correspondence Analysis, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA) to investigate the perceived role of various technological and governance concepts in shaping Sustainable Advantage (SA) within the literature. Correspondence Analysis revealed that terms such as «data,» «cloud,» «security,» «storage,» «learning,» and «transparency» exhibited higher contributions and inertias, suggesting they are central to the discourse. Factor loadings from EFA grouped constructs into distinct conceptual clusters -notably aligning «transparency,» «accountability,» and «sustainability» under a common factor-, which reflects the increasing emphasis on responsible technology deployment and ethical governance in data-centric environments.

The CFA-based structural models tested three mediators -Information Systems (IS), Machine Learning (ML), and Blockchain Technologies (BC- to explore how these constructs mediate the impact of Big Data Processing Technologies (BDPT) on Sustainable Auditing. Across all models, the direct effects of BDPT on SA were negative and statistically insignificant, suggesting that the solely availability of big data technologies does not translate into sustainable outcomes, unless it is accompanied by other organizational or technological capabilities. Among the mediators, only Blockchain showed significant positive direct effects on SA, result that is consistent with the current literature, which perceives them as enablers of long-term competitive advantages.

However, none of the indirect or total effects were statistically significant across the models, implying that the mediating roles of IS, ML, BC, or AC between BDPT and SA are either weak or inconsistent. These findings suggest that even though construct like Blockchain is seen as valuable for achieving sustainability goals, its interaction with big data technologies may not be straightforward or well-articulated in existing literature. Overall, the results of this study highlight a growing acknowledgment of the ethical, governance, and technological innovation dimensions in sustainability discussions, but also reflect a need for a stronger conceptual integration and empirical validation in future research.

6. Policy Implications and Future Research Directions

One key policy proposal is the promotion of governance frameworks that embed transparency and accountability, particularly within public sector institutions. Governments should mandate standardized protocols to ensure that organizations deploying data-driven technologies adhere to principles of ethical governance (Khan, 2024; Julien & Owusu-Berko, 2025). For instance, introducing «Sustainable Data Use Audits» that require institutions to report on the alignment of their data practices to environmental and social responsibility goals. While public institutions should focus on legal enforcement and oversight, private actors could be encouraged to adopt voluntary codes of conduct aligned with Environmental, Social, and Governance (ESG) models (Ong et al., 2025). This differentiation allows for a dual approach: a combination of government regulation and industry self-regulation, which will foster public trust.

A second recommendation involves the definition of incentives to promote the adoption of proven technologies -such as blockchain- to enhance traceability and verifiability, particularly within the private sector. Companies can integrate blockchain systems to ensure reliable audit trails, and credible sustainability certifications (Balaji, 2025; Thanasi-Boçe & Hoxha, 2025). To support this transition, governments might offer tax incentives or grants to firms investing in green or responsible technology infrastructures. Moreover, regulatory sandboxes could be established to pilot and evaluate blockchain innovations safely and effectively. While the public sector creates the enabling environment, private companies should focus on developing interoperability standards and upskilling their workforce to ensure seamless integration of blockchain technologies across sectors (Archana, 2025).

Another critical area for public-private joint action is the responsible use of big data and artificial intelligence (AI) in decision-making. Both sectors should implement governance structures that assess the social, ethical, and environmental implications of advanced analytics and machine learning. Establishing industry-wide certifications—similar to ISO standards—can help ensuring data quality, fairness, and alignment with sustainability principles (Mazé et al., 2016). Public institutions, for their part, should incorporate AI sustainability checklists into their procurement policies, thereby setting the standards for responsible technology use. In parallel, private enterprises could set up internal ethics review boards to monitor the societal impacts of their data-driven strategies.

Future research should focus on developing robust, interdisciplinary frameworks that integrate sustainability principles into the design, deployment, and governance of emerging technologies. There is a need to explore how to align data-intensive innovations—e.g., AI, IoT, and blockchain—with environmental and social goals, without compromising efficiency or scalability. Additionally, research should examine the role of cultural, regional, and institutional contexts in shaping ethical technology adoption, particularly in low-resource settings. Advancing such research will require collaboration among technologists, social scientists, environmental experts, and policymakers to ensure that digital transformation contributes positively to inclusive and sustainable development.

Referencias

Abreu, P. W., Aparicio, M., & Costa, C. J. (2018). Blockchain technology in the auditing environment. 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), 1–6. https://doi.org/10.23919/cisti.2018.8399460

Archana, T. (2025). Artificial Intelligence (AI) and Digital Competencies in the Public Sector. Digital Competency Development for Public Officials, 95–120. https://doi.org/10.4018/979-8-3693-6547-2.ch005

Aswar, K., Ermawati, & Julianto, W. (2021). Implementation of accrual accounting by the Indonesian central government: An investigation of social factors. Public and Municipal Finance, 10(1), 151–163. https://doi.org/10.21511/pmf.10(1).2021.12

Bai, C., Zhou, H., & Sarkis, J. (2023). Evaluating Industry 4.0 technology and sustainable development goals – a social perspective. International Journal of Production Research, 61(23), 8094–8114. https://doi.org/10.1080/00207543.2022.2164375

Balaji, K. (2025). Harnessing AI and Blockchain in Sustainability Assurance. Navigating Trust in Sustainability Reporting and Assurance, 215–242. https://doi.org/10.4018/979-8-3373-0117-4.ch007

Beh, E. J., & Lombardo, R. (Eds.). (2014). Correspondence Analysis. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9781118762875

Benbouzid, B. (2019). To predict and to manage. Predictive policing in the United States. Big Data & Society, 6(1), 205395171986170. https://doi.org/10.1177/2053951719861703

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, 99–113. https://doi.org/10.1016/j.techfore.2017.07.027

Cao, Q. (Ray), Chen, A. N. K., Ewing, B. T., & Thompson, M. A. (2021). Evaluating Information System Success and Impact on Sustainability Practices: A Survey and a Case Study of Regional Mesonet Information Systems. Sustainability, 13(13), 7260. https://doi.org/10.3390/su13137260

Demirel, E., & Eskin, İ. (2020). Investigation of the Effects of Environment on Financial Reporting. Ethics and Sustainability in Accounting and Finance, Volume II, 19–38. https://doi.org/10.1007/978-981-15-1928-4_2

Dovgal, V., & Kuizheva, S. (2022). Using Big Data Technology to Protect the Environment. Russian Journal of Earth Sciences, 1–5. https://doi.org/10.2205/2022es01si02

Ghobakhloo, M., Iranmanesh, M., Mubarik, M. S., Mubarak, M. F., Amran, A., & Khanfar, A. A. A. (2023). Blockchain technology as an enabler for sustainable business ecosystems: A comprehensive roadmap for socioenvironmental and economic sustainability. Business Strategy & Development, 7(1). Portico. https://doi.org/10.1002/bsd2.319

Harber, M., Maroun, W., & de Ricquebourg, A. D. (2023). Audit firm executives under pressure: A discursive analysis of legitimisation and resistance to reform. Critical Perspectives on Accounting, 97, 102580. https://doi.org/10.1016/j.cpa.2023.102580

Jayasuriya, D. D., & Sims, A. (2022). From the abacus to enterprise resource planning: is blockchain the next big accounting tool? Accounting, Auditing & Accountability Journal, 36(1), 24–62. https://doi.org/10.1108/aaaj-08-2020-4718

Julien, K. B., & Owusu-Berko, L. (2025). Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications. World Journal of Advanced Research and Reviews, 25(2), 1746–1763. https://doi.org/10.30574/wjarr.2025.25.2.0571

Khan, A. (2024). Data Ethics in the Age of AI. https://doi.org/10.1515/9781501522635

Lamboglia, R., & Mancini, D. (2021). The relationship between auditors' human capital attributes and the assessment of the control environment. *Journal of management and governance*, *25*(4), 1211-1239. https://doi.org/10.1007/s10997-020-09536-8

Li, B. (2021). Use of Social Audit Network Intelligent System to Promote High-Quality Economic Development in the Era of Big Data. In International Conference on Frontier Computing (pp. 205-212). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-8052-6_23

Ma, R. (2023). Construction of a social audit platform based on big data for «industry price, quality and credit.» Applied Mathematics and Nonlinear Sciences, 8(2), 1339–1354. https://doi.org/10.2478/amns.2023.1.00039

Mazé, A., Aït-Aïssa, M., Mayer, S., & Verjux, N. (2016). Third-Party Certifications and the Role of Auditing Policies in Sustainability. Organization & Environment, 29(3), 308–331. https://doi.org/10.1177/1086026615628034

Montero, A. G., & Le Blanc, D. (2019). The Role of External Audits in Enhancing Transparency and Accountability for the Sustainable Development Goals. (2019). UN Department of Economic and Social Affairs (DESA) Working Papers. https://doi.org/10.18356/3fe94447-en

Musunuru, K. (2024a). Easytm: An Easy Way to Text Mining in R. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4890734

Musunuru, K. (2024b). Textmining: A Python Script for Text Mining and Analysis. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4891783

Norris, M., & Lecavalier, L. (2009). Evaluating the Use of Exploratory Factor Analysis in Developmental Disability Psychological Research. Journal of Autism and Developmental Disorders, 40(1), 8–20. https://doi.org/10.1007/s10803-009-0816-2

Nwachukwu, C. E., Usman, T. O., Akhor, S. O., & Oladipupo, A. O. (2021). Auditing in the New Age of Industry 4.0. International Journal of Business Strategy and Automation, 2(1), 17–28. https://doi.org/10.4018/ijbsa.20210101.oa2

Oberst, A., & Gheorghita, M. (2022). Consolidate the reputation and increase the credibility of the company through the implementation of the social audit. Competitiveness and Sustainable Development, 104–110. https://doi.org/10.52326/csd2022.18

Ong, J. H., Khatibi, A., Mohd Talib, Z., & George, R. A. (2025). Ethical leadership in environmental, social and governance (ESG) adoption for Malaysian micro, small and medium enterprises (MSMEs). International Journal of Ethics and Systems. https://doi.org/10.1108/ijoes-08-2024-0266

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rahmadhani, S., Lim, J., & Santikawati, S. (2023). Analisis Praktik Audit Big Data Environtment di Indonesia. SINOMIKA Journal: Publikasi Ilmiah Bidang Ekonomi Dan Akuntansi, 1(5), 1135–1146. https://doi.org/10.54443/sinomika.v1i5.587

Revelle, W. (2020). How to: Use the psych package for factor analysis and data reduction. Northwestern University, Department of Psychology: Evanston, IL, USA.

Shahib, H. M., Sukoharsono, E. G., Achsin, M., & Prihatiningtias, Y. W. (2020). Developing Local Government's Socioenvironmental Accountability: Insights from Indonesian Socioenvironmental NGOs' Annual Reports. Environmentalism and NGO Accountability, 27–54. https://doi.org/10.1108/s1479-359820200000009003

Sucena, E., & Marinho, M. M. de O. (2019). Environmental disclosure analysis of sustainability reports the brazilian and international brewing industry based on Global Reporting Initiative - GRI. Gestão & Produção, 26(3). https://doi.org/10.1590/0104-530x3120

Sysoieva, I., Pozniakovska, N., Mikluha, O., Pukas, A., & Roleders, V. (2023). Social audit as a tool of civil society aimed at ensuring the sustainability. IOP Conference Series: Earth and Environmental Science, 1126(1), 012031. https://doi.org/10.1088/1755-1315/1126/1/012031

Tavares, M. C., & Azevedo, G. (2022). Contribution of Industry 4.0 Technologies to Social Responsibility and Sustainability. 2022 17th Iberian Conference on Information Systems and Technologies (CISTI). https://doi.org/10.23919/cisti54924.2022.9820334

Thanasi-Boçe, M., & Hoxha, J. (2025). Blockchain for Sustainable Development: A Systematic Review. Sustainability, 17(11), 4848. https://doi.org/10.3390/su17114848

Troshani, I., Doolin, B., Fradley, K., & Rampersad, G. (2022). Information systems, sociomaterial practices and the emergence of environmental management infrastructures. Australasian Journal of Information Systems, 26. https://doi.org/10.3127/ajis.v26i0.3829

Vinšalek-Stipić, V. (2022). The concept and importance of social audit for sustainable community development. BH Ekonomski Forum, 16(1), 151–165. https://doi.org/10.5937/bhekofor2201151v

Wang, Y. (2022). The Study Based on Intelligent Big Data Technology for Water Resources Audit. Scientific Programming, 2022, 1–9. https://doi.org/10.1155/2022/1188402

Wong, C. W. Y., Wong, C. Y., Boon-itt, S., & Tang, A. K. Y. (2021). Strategies for Building Environmental Transparency and Accountability. Sustainability, 13(16), 9116. https://doi.org/10.3390/su13169116

Yusoff, Y. H., Johari, A. S., Mohd Rahmatullah, D. A., Zainal, N. A., Tajuddin, N. A., & Thilaiampalam, N. T. S. (2023). Industry Revolution 4.0: Rapid Growth of Technology May Affect Job Security in Auditing Profession: A Concept Paper. International Journal of Academic Research in Business and Social Sciences, 13(3). https://doi.org/10.6007/ijarbss/v13-i3/16546

Zhang, K. (2023). Application of computer aided audit system and innovation research of enterprise financial audit under the background of big data. Advances in Engineering Technology Research, 5(1), 484. https://doi.org/10.56028/aetr.5.1.484.2023

Disclosure Statement
This is a study solely done by Dr. Kamakshaiah Musunuru, who has no relevant financial or non-financial competing interests with any other individual or organization.

Dr. Kamakshaiah Musunuru

Dr. M. Kamakshaiah is an academic in data science and analytics, and an open-source software evangelist. He has over two decades of teaching experience. Teaching Management Information Systems (MIS) and data science is his passion. He enjoys developing software applications for various data processing needs, such as machine learning, blockchain, big data, and IoT-related business practices. He has authored several books. Visit https://github.com/Kamakshaiah to explore his software applications.

Correo: kmusunur@gitam.edu

Revista Kawsaypacha: Sociedad y Medio Ambiente.
N° 15 enero – junio 2025. E-ISSN: 2709 – 3689

How to cite: Musunuru, K. (2025). Emerging Technologies for Socioenvironmental Auditing: Identification of Factors, Challenges and Technologies Using Text Mining and Analysis. Revista Kawsaypacha: Sociedad Y Medio Ambiente, (15), A-004. https://doi.org/10.18800/kawsaypacha.202501.A004