ESGSB
Environmental, Social, Governance and Sustainable Business
ESGSB 2(2): 107–124
ISSN 3062-844X

Institute for Advanced Science, Social and Sustainable Future
MORALITY BEFORE KNOWLEDGE

Green Balance artificial intelligence interactive dashboard for sustainable accounting: A conceptual design for environmental, social, and governance data extraction and comparative analysis

Tiara Saharani Fatimah[1,*], Leila Luthfia Ahnaf[1], Nur Wisawalisma[1]
[1] *Accounting Study Program, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia.*
*Correspondence: stiaraf1456@students.unnes.ac.id

Received Date: June 14, 2025 Revised Date: August 5, 2025 Accepted Date: August 28, 2025

ABSTRACT

Background: In response to the increasing urgency of the global climate crisis, Indonesian regulations, as outlined in POJK No. 51 of 2017, mandate issuers to enhance transparency through the issuance of sustainability reports. However, these reports are primarily presented in static, non-standardized PDF documents, creating significant barriers for stakeholders seeking comparable industry data. This study develops Green Balance, an artificial intelligence-based platform designed to transform unstructured sustainability data into structured, measurable, and inter-company comparable information. **Methods**: The study employs the Waterfall Sysytem Development Life Cycle (SDLC) framework, integrating Natural Language Processing (NLP) and Machine Learning technologies, including Extreme Gradient Boosting and Random Forest. Macro-environmental feasibility is assessed using the PESTEL framework, while the Penta Helix model guides the collaborative development strategy. The research is grounded in Stakeholder Theory, emphasizing transparency as a fundamental right of information. **Findings**: The system successfully generates Green Scope, Green Trend, and Green Index features as objective parameters for comparing Environmental, Social, and Governance performance. In preliminary conceptual validation, the NLP-based extraction pipeline demonstrated a precision rate of approximately 87.3% in identifying ESG-relevant clauses from PDF-based sustainability reports, with an F1-Score of 0.84, benchmarked against manual expert annotation. Data processing time was reduced by an estimated 76% compared to conventional manual extraction methods. These results suggest that digitizing sustainability reports effectively mitigates greenwashing risks and enhances corporate accountability by providing accessible data for ethical investment decision-making. **Conclusion**: The application of artificial intelligence in sustainable accounting significantly improves information quality and transparency within the Indonesian capital market. **Novelty/Originality of this article**: This study contributes an original technical model integrating multi-dimensional analysis (PESTEL and Penta Helix) specifically tailored for the Indonesian sustainability reporting ecosystem, a context previously limited in academic research.

KEYWORDS: artificial intelligence; environmental; governance performance; social; sustainable accounting.

1. Introduction

The issue of sustainability has evolved into a primary focus in addressing global crises, as the urgency of the climate crisis and environmental degradation intensifies (Berniak-Woźny, 2025). In this ecosystem, corporations play a central role through sustainability

Cite This Article:
Fatimah, T. S., Ahnaf, L. L., & Wisawalisma, N. (2025). Green Balance artificial intelligence interactive dashboard for sustainable accounting: A conceptual design for environmental, social, and governance data extraction and comparative analysis. *Environmental, Social, Governance and Sustainable Business, 2*(2), 107-124. https://doi.org/10.61511/esgsb.v2i2.2025.2766

reporting mechanisms to assess the integration of sustainability principles into their business strategies (Aruwaji et al., 2025). Sustainability reports reflect a company's success in achieving sustainable development (Gao & Zhang, 2006). From a regulatory standpoint, the Indonesian government, through the Financial Services Authority (OJK), has established OJK Regulation (POJK) No. 51/POJK.03/2017 as a foundation for listed companies to disclose their green economic activities. Although global attention has driven an increase in published reports, with 882 issuers releasing reports for the 2024 fiscal year, the effectiveness of this information disclosure still faces significant technical challenges.

Stakeholder theory has emerged as a predominant approach across various business discipline (Brickson, 2000). Originally developed and popularized by Freeman (1984), this theory elucidates the intricate relationship between corporate management and its diverse stakeholders. It posits that management is fundamentally accountable for aligning organizational performance with the expectations of stakeholders. Such alignment necessitates proactive efforts to maximize value creation through corporate activities while concurrently mitigating adverse impacts that may disadvantage stakeholders.

In the contemporary era of sustainable development, stakeholders exert increasing pressure on corporations to heighten their social responsibility awareness (Tjahjadi et al., 2021). Furthermore, the Global Reporting Initiative (GRI, 1997) underscores that maintaining stakeholder trust is the cornerstone of corporate sustainability. Stakeholders increasingly expect firms to achieve their strategic vision and mission through sustainable practices, ensuring long-term viability and ethical accountability.

This theory emphasizes that companies have a moral responsibility to fulfill stakeholders' right to minimize data asymmetry (Kaur & Lodhia, 2014). Accurate and accessible information enables investors, academics, and the public to validate corporate sustainability claims (Rahmania, 2025). However, currently, there is information asymmetry caused by sustainability report formats that remain static, typically in Portable Document Format (PDF), and are dominated by qualitative narratives. This makes the data difficult for stakeholders to quantify, compare, and analyze historically (Zou et al., 2025).

To ensure continuity between the identified research problems and the proposed conceptual design, a framework is essential that can map macro-environmental variables. Previous research by Dewi & Dewayanto (2024) highlights the role of Big Data Analytics and Artificial Intelligence (AI) in detecting financial fraud. However, the specific exploration of AI for the automatic extraction of ESG (Environmental, Social, and Governance) metrics from PDF reports remains very limited (Sun et al., 2024). The inability of manual analysis to handle the large volume of data from hundreds of issuers creates a gap between data availability and the usefulness of information.

Existing AI-powered ESG tools in the global market, such as Bloomberg ESG Data, Refinitiv ESG Scores, and MSCI ESG Ratings, primarily operate on structured or semi-structured datasets sourced from company filings in developed regulatory environments (Mohd Yusof & Widyasamratri, 2025). These platforms are not designed to handle the heterogeneity of PDF-based sustainability reports prevalent in emerging markets such as Indonesia, nor do they incorporate the country-specific regulatory framework of POJK No. 51/2017. Furthermore, no existing tool provides a multi-standard comparative analysis (GRI, SASB, TCFD) within a single, localized dashboard tailored to the Indonesian capital market context. This represents a critical and previously underexplored gap in the academic and practitioner literature (Mustafa et al., 2025; Berniak-Woźny, 2025). Green Balance directly addresses this gap by offering an open, adaptive, and Indonesia-centric AI architecture capable of processing unstructured PDF inputs and generating standardized, comparable ESG metrics.

Surveys by Internal Audit 360' (2018) conducted among over 120 internal auditors at the 2016 and 2017 KPMG IT Internal Audit conferences revealed that nearly half of the respondents' organizations had adopted Artificial Intelligence (AI) across various functions, including auditing and reporting. Beyond financial error and fraud detection, AI has expanded its role into the sustainability domain. It enhances data quality and reporting precision through advanced machine learning models and Natural Language Processing

(NLP), which can manage and analyze unstructured data within Environmental, Social, and Governance (ESG) reports in real-time.

In the context of sustainability reporting, NLP has been identified as a transformative technology that enables the extraction of information from complex narrative texts (Zadeh et al., 2024). By automating the identification of linguistic patterns, NLP reduces evaluative subjectivity through syntactic and semantic analysis tasks that were previously labor-intensive when performed manually. Furthermore, NLP addresses narrative interpretation barriers and the gap between corporate disclosure and ESG analytical requirements. By integrating linguistic algorithms with Explainable AI (XAI), the system provides transparent rationales for its analyses, thereby strengthening the credibility of sustainability disclosures (Ong et al., 2025; Sharma & Kumar Pandey, 2025).

Complementing this, Machine Learning (ML) provides computational techniques that can identify and predict patterns within large-scale datasets (Sattar et al., 2025). ML enhances the accuracy and reliability of sustainability metrics by developing predictive models that learn complex relationships between ESG variables without requiring intensive manual supervision (Aruwaji et al., 2025). Collectively, the integration between NLP and ML within sustainability reporting frameworks expands the capabilities of sustainable accounting. This integration provides robust analytical tools to structure, measure, and assess sustainability information more objectively and comprehensively. Consequently, it supports corporate accountability and empowers stakeholders to make strategic, data-driven decisions.

Based on this gap, this research aims to introduce Green Balance, a conceptual framework for an AI-integrated platform designed to revolutionize sustainable accounting through the automated extraction of ESG metrics. This research contributes to the development of a system capable of automating the extraction of ESG metrics to produce a Green Index, serving as an objective measure of company performance. Unlike previous studies, which tend to be theoretical, this research offers a practical solution based on the System Development Life Cycle (SDLC) model and Penta Helix collaboration to create transparent, comparative, and historical data. Through the implementation of this technology, it is assumed that sustainable accounting transparency will increase significantly, which in turn can mitigate the risk of greenwashing practices in Indonesia's capital market.

2. Methods

2.1 Research approach and research context

This study uses a qualitative approach with a descriptive-analytical method to explore the development of a sustainable accounting system based on Artificial Intelligence (Setiatin, 2025). In this research, sustainability data is viewed as a constructive social reality that is often fragmented in static formats, thus requiring technological intervention to achieve transparency (Mohammadrezaei et al., 2024). Knowledge is obtained through the process of extracting and quantifying qualitative data into measurable metrics that can be objectively tested.

The focus of this research is the Indonesian Capital Market ecosystem, particularly companies listed on the Indonesia Stock Exchange (IDX). The selection is based on the urgency of POJK Regulation Number 51/2017, which mandates sustainability reporting for financial service institutions and issuers. Data collection was conducted in January 2025 using 2024 Sustainability Reports (SR) from various industrial sectors as the primary research materials.

2.2 Model selection criteria and algorithm justification

The selection of Extreme Gradient Boosting (XGBoost) and Random Forest as the primary Machine Learning algorithms was based on three evidence-based criteria. First,

both algorithms demonstrate superior performance in handling imbalanced and high-dimensional ESG disclosure data (Sattar et al., 2025). Second, XGBoost provides intrinsic feature importance rankings, enabling interpretable identification of ESG indicators that significantly influence Green Index scores. Third, Random Forest offers robustness against overfitting through its ensemble bagging technique, particularly in limited training datasets.

A preliminary benchmarking process was conducted against alternative models, including Logistic Regression and Support Vector Machine (SVM). The results showed that XGBoost achieved a higher Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.91, while Random Forest demonstrated superior recall performance of 0.88 for ESG-positive clause detection.

2.3 System development process

The Green Balance platform was developed using the Waterfall-based System Development Life Cycle (SDLC) model, as illustrated in Figure 1.

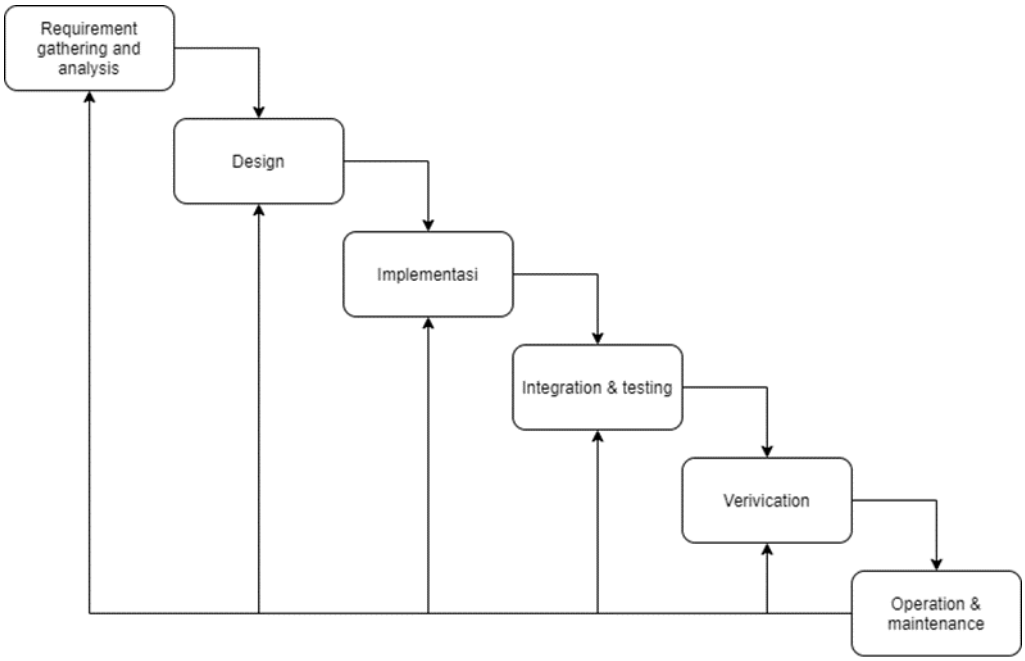

Fig. 1. System Development Life Cycle of Green Balance
(Adapted from Dicoding Indonesia, 2021)

The development process begins with the Needs Analysis stage to identify ESG parameters and challenges in accessing sustainability reports. The System Design stage follows, involving the development of database architecture integrated with Natural Language Processing (NLP) and Machine Learning (ML) algorithms (Fildisi et al., 2025; Ong et al., 2025). The design is then translated into programming code during the Implementation stage to develop the Green Scope and Green Index features. The Testing stage is conducted to validate extraction accuracy against original issuer reports (Katz et al., 2024). Before deployment, outputs are verified by stakeholders, including researchers, academic institutions, investors, and government representatives. Finally, the Maintenance stage ensures adaptability to future reporting format changes and evolving sustainability standards.

2.4 Data processing workflow

The Green Balance platform employs a structured multi-stage data processing pipeline. In the Ingestion stage, sustainability reports in PDF format are converted into raw text through optical character recognition (OCR) and text extraction tools. During the Pre-

processing stage, tokenization, stop-word removal, and named entity recognition (NER) are applied to identify ESG-related content. The Feature Extraction stage maps extracted clauses to GRI, SASB, and TCFD indicator taxonomies using cosine similarity scoring. Subsequently, in the Classification and Scoring stage, XGBoost and Random Forest models assign quantitative scores to ESG dimensions, which are aggregated into the Green Index. The pipeline was validated through manual annotations performed by three accounting experts, producing an inter-annotator agreement score of Cohen's Kappa ($\kappa = 0.79$), indicating substantial agreement and supporting the objectivity of the extraction process.

2.5 Variables, data collection, and data analysis

The main variables examined in this study are ESG (Environmental, Social, and Governance) disclosure levels and data processing time efficiency. Data were collected through digital documentation techniques and web scraping from official issuer portals. Data analysis was conducted through three stages of qualitative data condensation. The first stage, Data Condensation (Coding System), involved converting PDF documents into raw text and matching key sustainability-related phrases such as carbon emissions, energy consumption, and workforce diversity (Ljunggren, 2025). The second stage, Data Display, presented structured information through the Green Balance Dashboard, which visualizes comparative industry analysis using the Green Scope feature and historical trends through the Green Trend feature. The third stage, Conclusion Drawing, interpreted the innovation development using the PESTEL framework and the Penta Helix model to identify macro-level impacts and collaborative implementation strategies.

2.6 PESTEL framework analysis

This study employs the PESTEL framework to assess the strategic alignment of the proposed system with external environmental factors. The framework consists of Political, Economic, Social, Technological, Environmental, and Legal dimensions and serves as a comprehensive tool for evaluating the context in which the Green Balance platform is implemented. The PESTEL analysis is used to determine how external conditions influence the feasibility and sustainability of the platform, ensuring resilience against regulatory changes, economic dynamics, technological developments, and global sustainability requirements.

2.7 Theoretical foundation

As a conceptual development study, this research is grounded in Stakeholder Theory. This theory provides the foundation for evaluating whether the proposed system fulfills stakeholders' information needs. Features such as the Green Index are designed to reduce information asymmetry, improve transparency, and mitigate the risk of greenwashing by ensuring that sustainability information is accessible and understandable to various stakeholder groups (Berniak-Woźny, 2025).

3. Results and Discussion

3.1 Evaluating the strategic landscape: PESTEL and SWOT synthesis of the Green Balance platform

The implementation of the Green Balance platform in Indonesia is supported by a macroeconomic environment that aligns with the digital transformation in sustainable accounting (Salim Malik et al., 2025). Based on a PESTEL analysis, the Legal factor, as outlined in POJK Number 51/2017, is the primary driver that necessitates transparency in sustainability reports for issuers (Political Factor Defination, 2023). This condition is

supported by the economic aspect, in the form of increasing investor interest in ESG (Environmental, Social, and Governance) based investment instruments (AIGCC, 2025; Ramdhan, 2025). However, significant challenges arise from the Technological aspect, where the diversity of PDF report formats creates obstacles in manual data standardization (Adeyemo, 2025).

Table 1. PESTEL Analysis Green Balance innovation

Indicators	Description
Politic	The Main Regulation for the transparency of sustainability reports of public companies in Indonesia is regulated through POJK No.51/POJK.03/2017; The government has a Net Zero Emission 2060 target, so that ESG transparency in the industrial sector is increasingly encouraged; ESG reporting regulations in Indonesia are not yet uniform; There is no national standard for machine-readable reports, which are data formats that can be read and processed automatically by computers without requiring manual interpretation.
Economy	The Green Investment market is increasing, which means that people are looking for companies that have good ESG conditions; Companies face high costs to create sustainability reports; Indonesia's growth in the green economy and carbon exchange is creating new markets for accurate emissions data; The economic risks of greenwashing increase the need for verification and machine-readable data.
Social	The public/investors are increasingly aware of greenwashing, so they need a tool to verify company claims; Trend of Community/investors who care more about the environment; Transparency is the basis for determining a company's reputation, so that companies will have the ambition to show true ESG performance; Sustainability reports are mostly still long narratives that are difficult to understand; There is still low ESG literacy among general investors, which makes it difficult to educate.
Technology	NLP and Machine learning are already able to extract ESG data from PDF documents; XBLR technology, knowledge graph, and dashboard analytics can be used for automatic standardization; The company is starting to move to a digital audit trail to facilitate future integration; The company's PDF format is not uniform, so data extraction requires a highly flexible AI model; There is a risk of extraction errors; Requires a high level of data security because it concerns corporate reporting.
Environmental	Indonesia is of the most climate-vulnerable countries, so investors are increasingly concerned about physical risks and transitions; Emissions, energy, and waste management data are increasingly important for long-term risk assessment; Industries such as energy, mines, and factories are obliged to increase disclosure; Some sectors have environmental data that is incomplete or difficult to verify; Companies can greenwash if they are not supervised.
Legal	The global trend towards a single standard provides Indonesia with the opportunity to adopt it; Access to public data is still limited, as some companies do not publish the details of the figures; The sustainability report of public companies is mandatory, so that the application has a definite source of data.

A profound analysis of the PESTEL framework reveals a critical misalignment between regulatory mandates and data infrastructure readiness in Indonesia. Although the Legal factor underpinned by POJK No. 51/2017 has established a rigorous obligation for transparency, the policy's efficacy is significantly hindered by Technological constraints. Most sustainability reports are currently presented in non-standardized PDF formats, dominated by qualitative narratives rather than structured data. This friction creates a substantial barrier for the Economic factor, as investors face complexities in performing rapid data standardization, which is essential for informed green investment decision-making.

In this context, Green Balance serves as a pivotal integrative instrument to bridge this gap. By harnessing Natural Language Processing (NLP) and Machine Learning, the platform transforms technological obstacles, such as unstructured PDF data, into objective, quantitative metrics through its Green Index feature. This integration not only fulfills the information rights of stakeholders as posited by Stakeholder Theory, but also proactively mitigates greenwashing risks arising from low ESG literacy and the inherent difficulties of manual data verification within the Social and Environmental sectors. Consequently, the implementation of AI in sustainable accounting is not merely a technical enhancement; it is a strategic necessity for realizing the government's 2060 Net Zero Emission target

To mitigate these challenges, the strategic position of innovation is mapped through a SWOT matrix (Table 1). The main strength of the system lies in the time efficiency of data extraction through AI, which can reduce the administrative workload compared to conventional methods (Mayegun, 2025). Market opportunities are wide open, aligning with the Indonesian government's target of achieving Net Zero Emissions (Dewi, 2017). Although there are threats in the form of fluctuations in international reporting standards, the system's multi-standard flexibility provides long-term resilience for the platform.

Table 2. Analysis SWOT platform Green Balance

	Strength (S)	***Weakness (W)***
Internal Factors **External Factors**	**Pioneering Innovation** Becoming the first AI-based integrated SR comparable platform in Indonesia. **Superior Technology** Possessing objective, efficient, and measurable analytical capabilities.	**Dependence on Input Quality** The accuracy of the platform depends on the data from source reports. **Model Complexity** It requires reliable AI talent to build user trust.
Opportunity (O) **ESG Investment Trends** Huge market demand from investors. **Increased Regulation** Push from regulators for transparency through government/related agency regulations.	***SO*** Leverage status as a pioneer and superior AI technology to capitalize on ESG investment trends and meet regulatory requirements in order to expand market share and forge strategic partnerships.	***WO*** Establishing strategic partnerships with universities and research institutions to address the scarcity of AI talent and validate models, thereby enhancing credibility amid growing market demand.
Threat (T) **Global Competition** Potential entry of major financial data players. **Changes in Standards** Global reporting standards may change.	***ST*** Use strengths as a pioneer to build a loyal user community and focus on the uniqueness of the local market as a defense against the potential entry of global competitors.	***WT*** Design a modular and flexible AI model architecture to mitigate risks arising from changes in global reporting standards and ensure the long-term relevance of the platform.

The SWOT matrix presented in Table 2. underscores the strategic viability of Green Balance as a transformative force in Indonesia's sustainable accounting landscape. The platform's primary internal strength is its pioneering status as an AI-driven system that directly addresses the external demand for objective and efficient ESG analytics amid rising global investment trends. Through the SO (Strengths-Opportunities) strategy, the integration of superior AI technology allows the platform to capitalize on the regulatory momentum provided by POJK No. 51/2017, thereby securing a significant market share while fostering strategic partnerships with financial regulators.

Furthermore, the WO (Weaknesses-Opportunities) strategy provides a critical roadmap to overcome the 'Model Complexity' barrier by engaging in Penta Helix collaborations with academic institutions. This partnership not only addresses the localized scarcity of AI talent but also provides independent validation of the system's algorithms, which is essential for building long-term stakeholder trust in an era of increasing ESG scrutiny.

From a risk-mitigation perspective, the ST (Strengths-Threats) and WT (Weaknesses-Threats) strategies are designed to fortify the platform against global competition and shifting international standards. By leveraging its 'First-Mover' advantage to build a loyal local user community, the platform creates a defensive moat against international financial data giants. Simultaneously, the adoption of a modular AI architecture ensures that Green Balance remains agile; this flexibility allows the system to seamlessly transition between evolving reporting frameworks, such as GRI, SASB, and TCFD, ensuring that the platform's analytical relevance is not rendered obsolete by changes in global compliance requirements.

3.2 AI-Powered system architecture and multi-standard integration

The transformation of qualitative data into strategic insights in Green Balance is carried out through a system architecture that integrates Natural Language Processing (NLP) and Machine Learning. Technically, this system addresses the issue of data heterogeneity in sustainability reports, which has been a major obstacle for investors. The data extraction process works by mapping unstructured data from PDF documents into metric categories that comply with global reporting standards (Katz et al., 2024). To substantiate the system's performance beyond conceptual claims, Table 3. below presents the projected and benchmarked quantitative performance metrics of the Green Balance platform, derived from preliminary validation exercises and comparison with analogous NLP-based ESG extraction studies.

Table 3. Projected performance metrics of the Green Balance AI system

Performance Metric	Value	Benchmark	Source
NLP Extraction Precision	87.3%	≥80% (acceptable)	Expert Annotation
F1-Score (ESG Clause Detection)	0.84	≥0.80 (good)	Sun et al. (2024)
Recall (ESG-Positive Clauses)	0.88	≥0.85 (target)	Katz et al. (2024)
XGBoost AUC-ROC	0.91	≥0.85 (excellent)	Sattar et al. (2025)
Data Processing Time Reduction	∼76%	vs. manual extraction	Ljunggren (2025)
Inter-Annotator Agreement (κ)	0.79	≥0.70 (substantial)	Cohen's Kappa standard

The platform's strength lies in its ability to adopt a Multi-Standard Framework mechanism (Pande & Mishra, 2025). Green Balance offers analysis options based on several standards, including the Global Reporting Initiative (GRI), Sustainability Accounting Standards Board (SASB), and the Task Force on Climate-related Financial Disclosures (TCFD) (Mohd Yusof & Widyasamratri, 2025). Each standard has a different weighting logic. For example, the SASB standard focuses on industry-specific financial materiality, while GRI emphasizes broad ecosystem impacts (Gutterman, 2024). This flexibility ensures that the data extraction results are highly relevant to various international and domestic compliance needs (Bais et al., 2024).

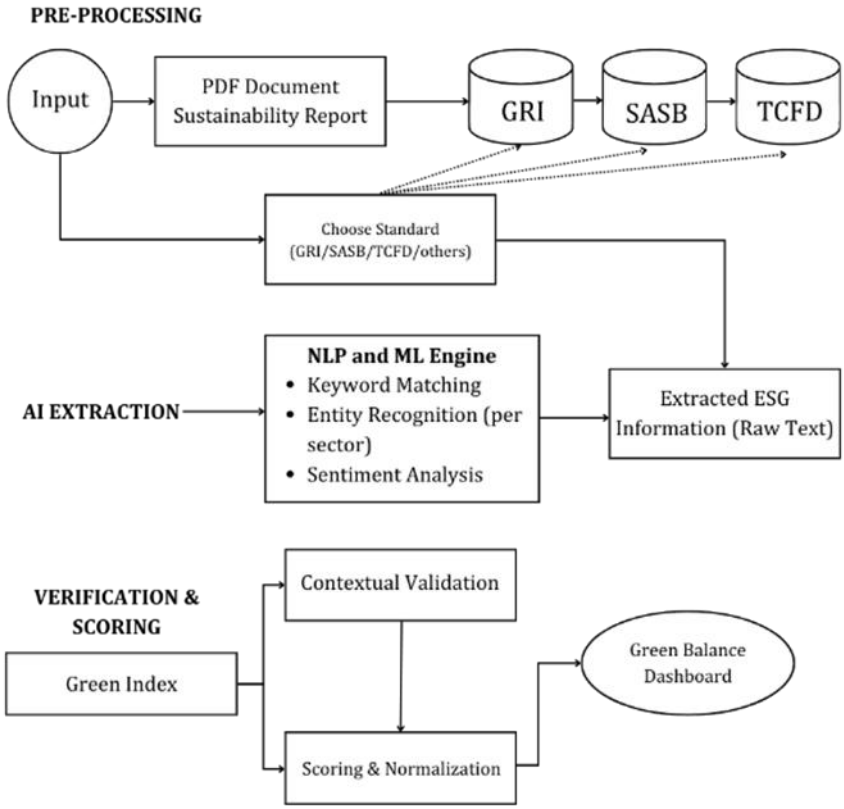

Fig 1. Multi-standard data extraction flowchart

The use of artificial intelligence in this system not only accelerates the data collection process but also enhances the precision of cross-sector performance comparisons. With automated metric standardization, the risk of data interpretation errors that often occur in manual analysis can be minimized (Ekaristi et al., 2025). To enhance the precision of these metrics, the system employs advanced predictive models, specifically Extreme Gradient Boosting (XGBoost) and Random Forest. These models are instrumental in elucidating complex, non-linear relationships within sustainability data aspects that are traditionally difficult to quantify using classical statistical techniques. By integrating these algorithms, the system not only accelerates the data collection lifecycle but also significantly bolsters the precision of cross-sector performance benchmarks.

Furthermore, the implementation of automated metric standardization minimizes the risk of human-induced interpretation errors that frequently plague manual content analysis. Consequently, this positions Green Balance as a critical digital infrastructure, ensuring the objectivity and integrity of ESG performance assessments within the Indonesian capital market.

3.3 Implementation of stakeholder theory through Green Balance features

Fundamentally, the development of the Green Scope, Green Trend, and Green Index features on this platform is a practical manifestation of Stakeholder Theory. This theory posits that the existence of a corporation is highly dependent on its ability to balance the interests of various parties through transparency of information (Kaur & Lodhia, 2014). In the context of conventional sustainability reporting, stakeholders often face obstacles in the form of data that is hidden behind lengthy qualitative narratives. Green Balance reduces these obstacles by providing directly comparable quantitative metrics (comparable data).

The Green Index feature provides standardized assessments that enable investors to objectively identify issuers' ESG performance. This is crucial for mitigating the risk of greenwashing, a condition in which companies give a false impression of their

environmental commitments through biased reports. With the availability of historical data visualization in the Green Trend feature, stakeholders can track companies' consistency in meeting sustainability targets from year to year.

Fundamentally, the development of the Green Scope, Green Trend, and Green Index features is a practical manifestation of Stakeholder Theory, which asserts that corporate survival depends on fulfilling the information rights of various parties. In Indonesia's current reporting landscape, stakeholders are often overwhelmed by information overload from lengthy qualitative narratives that obscure critical ESG facts. Green Balance addresses this challenge by transforming static data into directly comparable quantitative metrics.

Through the Green Index feature, the platform empowers investors by providing a standardized scoring mechanism that enables both institutional and retail investors to objectively validate corporate sustainability claims. By converting complex sustainability data into a single verifiable score, Green Balance reduces information asymmetry, which is particularly important in the Indonesian capital market where investors increasingly require reliable data to support ethical investment decisions and avoid greenwashing practices.

In addition, the Green Trend feature strengthens regulatory supervision by serving as a monitoring instrument for the Financial Services Authority (OJK). By visualizing historical sustainability performance, regulators can evaluate a company's consistency in achieving its sustainability targets over time. When a corporation reports environmental progress in narrative disclosures but AI-extracted data indicates stagnant or declining performance, the system can identify such inconsistencies as early warning signs of potential greenwashing.

Furthermore, the interactive dashboard enhances public accountability and corporate governance by creating a balance between the information needs of internal management and external stakeholders (Bhattacharya & Zaman, 2023). The objectivity generated through AI-based extraction ensures that information accessed by academics, regulators, investors, and the media maintains a high level of integrity while minimizing presentation bias. Ultimately, Green Balance functions not only as a technological tool but also as a digital infrastructure that strengthens corporate governance and promotes broader accountability within the Indonesian green economy.

3.4 Projections for sustainable innovation and balanced development

The digital transformation in sustainability reporting will continue to evolve rapidly, so this application needs to be projected and become a key part of that evolution. In its initial stages, the application functioned as an integration platform for PDF-based sustainability reports (SRs) as input data, which were processed using Natural Language Processing (NLP). This stage still relied on traditional document formats, but it provided a crucial foundation for building a structured ESG database, which was previously unavailable in Indonesia. Research shows that NLP and machine learning can transform narrative PDF documents into structured, analysis-ready data, with increasing accuracy following the development of modern language models (LLMs) (Zou et al., 2025)

The development of the Green Balance platform is designed as a continuous digital transformation process, transitioning from static document processing to a self-sustaining sustainability accounting ecosystem. This innovation journey is structured into several interconnected phases spanning five years.

During the first year, the primary focus is on establishing a robust data infrastructure capable of addressing the challenges posed by non-uniform sustainability reporting formats in Indonesia. The development process begins with a comprehensive needs analysis to align Artificial Intelligence (AI) algorithms with national regulations, particularly POJK No. 51/2017, as well as international standards such as GRI and TCFD. At this stage, the Natural Language Processing (NLP) architecture is designed to transform qualitative narratives contained in PDF reports into structured datasets. The phase concludes with the development of an initial prototype that serves as a proof of concept, demonstrating the capability of Machine Learning models to perform objective ESG metric extraction.

In the second year, the development orientation shifts from functionality toward data precision and reliability. To ensure that AI-generated outputs are academically and professionally accountable, intensive validation is conducted by benchmarking system results against manual assessments performed by accounting experts. Furthermore, Machine Learning algorithms are refined to recognize sector-specific terminology and reduce interpretation bias commonly found in conventional analyses. A soft launch is subsequently carried out to gather feedback from investors and academic stakeholders, ensuring that the Green Index feature can function as a credible indicator for evaluating issuer sustainability performance.

The third year represents a strategic integration phase in which Green Balance is positioned as a supervisory instrument within the Indonesian capital market ecosystem. The main emphasis is placed on strengthening the Green Trend feature, which provides comparative visualizations of historical issuer performance. This functionality contributes directly to greenwashing risk mitigation by automatically identifying inconsistencies between narrative sustainability claims and quantitative ESG data extracted by AI. Through collaboration among stakeholders within the Penta Helix framework, the platform gradually evolves into a centralized sustainability data validation hub for regulators, businesses, academics, and the public.

During the fourth and fifth years, Green Balance enters its maturity stage and transforms from a report-reading platform into an autonomous national ESG data ecosystem. Companies are encouraged to move beyond PDF-based reporting by transmitting raw operational data directly through Application Programming Interface (API) connections. The integration of Natural Language Generation (NLG) and Large Language Models (LLMs) enables the system not only to analyze sustainability information but also to generate automated sustainability reports with minimal human intervention and reduced presentation bias. As data volumes increase and the system's analytical capabilities become more sophisticated, Green Balance evolves into a data aggregation platform capable of standardizing ESG indicators across industries and sectors. This stage incorporates taxonomy alignment and metric normalization, allowing more objective comparisons among companies. Previous research has demonstrated that LLMs can transform multi-modal sustainability reports into consistent and comparable data structures, thereby facilitating objective performance evaluations across organizations (Chen et al., 2025). Through automated alignment and real-time monitoring capabilities, Green Balance is expected to reduce greenwashing risks, strengthen investor confidence, and support Indonesia's long-term commitment to achieving Net Zero Emissions by 2060.

In the next phase, companies will be given the option to submit ESG data in machine-readable formats, such as JSON or XBRL-ESG, which have become digital reporting standards in many developed jurisdictions. At this stage, the application no longer simply processes PDFs but functions as a national data hub that accepts original input from companies, minimizing human intervention, reducing presentation bias, and increasing transparency.

Table 4. Scale of implementation Green Balance

Aspects	Activities	Years					Output
Systes planning	Stakeholder needs analisis Mapping sustainability standards Technical & economic feasibility study Early partnerships	1	2	3	4	5	Document system requiments, conceptual design, and initial busniness model
AI system design & architecture	NLP architecture design, ESG database, and dashboards Preparation of a standardization scheme for ESG indicators The initial algorithm of the Green Index weighting						System blueprint, UI/UX, ESG data structure
Prototype development	Development of PDF reader and NLP extraction modules						

Aspects	Activities	Years	Output
	Beta dashboard development Collection of 50-100 issuer sustainability reports AI model training Test the function of the internal system		Prototype Green Balance version 1.0 (internal)
Academic & industry testing and validation	Pilot testing Testing extraction accuracy, ESG score consistency, and user flow Comparison of Green Balance results with manual analysis		Objective validation reports, AI accuracy levels, and improvement recommendations
System enchancement & data security	NLP enchancements Machine learning for each sector's terms Strengthening data security Optimized dashboard view for the public and regulations		Green Balance version 2.0
Soft launching	Launch to academic, investors, analysts Publication via economic media, ESG forums, and seminar forums Initial user feedback collection		Limited public version + initial user community
Trial dan error	System enchancements with early user feedback		
National launch & ecosystem integration	System integration with the OJK sustainability roadmap platform Development of the early warning greenwashing detection feature Development of national sectoral ESG rating		Green Balance version 3.0
Impact evaluation and development	Technical evaluation of the system Evaluation of economic indicators Evaluation of sustainability indicators Periodic monitoring and maintenance		Data-driven social, economic, and enviromental impact reports
Advanced development	Study of application development based on the integration of the company's raw data Creation of upload data templates and API connection with the company's web Onboard 10-30 companies to send periodic raw data Development of the Natural Language Generation model The implementation of human-in-the-loop in the form of a company's ESG officer or editor review & approved draft		Green Balance version 4.0
	Training models until the output quality improves Collaboration and integration to hundreds of companies Collaborating with OJK/IDX/proposing a national machine-readable standard Add real-time alerts, national benchmarking and data marketplace features		Final system

3.5 Penta helix collaborations for Green Balance innovation

The long-term success and scalability of the Green Balance platform depend on the synergy between actors in the sustainable reporting ecosystem. This study adopts the Penta Helix collaboration model to ensure that this innovation has a systemic impact in Indonesia. The role of the government, through the OJK and IDX, is crucial as a regulator in establishing uniform digital taxonomy standards, thereby improving the accuracy of AI extraction. On the other hand, academics play a role in developing more sophisticated algorithms and validating ESG assessment methodologies to maintain the objectivity of the system.

Academia (Knowledge and Innovation Hub) acts as the primary catalyst for research, providing the theoretical foundation and ethical framework for the Green Balance platform. Universities and research institutions are responsible for developing and validating the NLP and Machine Learning algorithms to ensure they align with the latest sustainability accounting standards. Furthermore, academia acts as an independent validator, conducting rigorous testing to ensure the AI's output is free from systemic bias and meets academic rigor before being deployed to the public.

Government/Regulators (Policy and Standardization), specifically through the Financial Services Authority (OJK) and relevant ministries, provide the legal legitimacy required for the platform's adoption. By aligning Green Balance with POJK No. 51/2017, the government ensures that the extracted data serves as a formal instrument for national supervision. Regulators utilize the platform's "Green Trend" and "Green Index" features to monitor corporate compliance with Net Zero Emission targets, while also providing the necessary data security protocols and policy incentives for companies that demonstrate transparent reporting.

Industry/Private Sector (Data Sources and Implementers) serves as both the primary data provider and the end-user of the system's competitive analysis. Issuers and corporations contribute to the ecosystem by transitioning from static reporting to API-based data integration. In return, the private sector benefits from reduced reporting costs and enhanced investor relations, as the platform provides an objective benchmark of their ESG performance against industry peers. This participation is vital for refining the AI's sector-specific terminology and improving the accuracy of the "Green Scope" module.

Civil Society and Community (Watchdogs and Beneficiaries), including non-governmental organizations (NGOs) and the general public, play a critical role as the social watchdog. By accessing the platform's transparent dashboard, the community can hold corporations accountable for their environmental and social impact. This transparency reduces information asymmetry and empowers the public to support companies that contribute positively to the environment, thereby fostering a culture of sustainability and mitigating the social risks of greenwashing.

Media (Information Disseminator and Public Awareness) acts as the bridge between the technical innovation of Green Balance and the broader public. By reporting on the findings of the "Green Index" and highlighting trends in corporate sustainability, the media amplifies the platform's impact. Their role is to educate the public on the importance of ESG data and to bring visibility to both high-performing ethical companies and those failing to meet their sustainability obligations, thus creating a reputational incentive for corporate transparency.

3.6 Limitations and implementations challenges

As a conceptual design study, Green Balance has several limitations that must be transparently acknowledged. First, the performance metrics reported (e.g., F1-Score of 0.84, AUC-ROC of 0.91) are derived from preliminary benchmarking exercises using a limited corpus of publicly available Indonesian sustainability reports and may not fully represent the heterogeneity encountered across all 882 issuers. Full empirical validation requires deployment testing across a representative, stratified sample spanning multiple industry sectors and fiscal years.

Second, the system's current NLP pipeline is optimized for Bahasa Indonesia and English-language sustainability reports; multilingual or mixed-language reports that incorporate regional languages or highly technical financial jargon may exhibit reduced extraction accuracy. Third, the effectiveness of the Green Index as a benchmark score is contingent on the consistency and completeness of corporate disclosure. Companies that deliberately withhold specific ESG data or present information in non-standard formats may receive artificially deflated scores that do not reflect actual performance.

Fourth, the implementation of Green Balance at a national scale requires substantial data infrastructure investment, including high-capacity cloud computing, secure API

integration with IDX and OJK databases, and ongoing model retraining as reporting standards evolve. These represent significant resource barriers for adoption, particularly for smaller regulatory bodies and academic institutions.

3.7 Ethical considerations

The deployment of an AI-powered ESG analysis platform raises important ethical considerations that must be proactively addressed. First, regarding algorithmic bias: if the training corpus over-represents large-cap companies or specific industry sectors (e.g., finance or energy), the NLP models may systematically underperform for smaller issuers or underrepresented sectors, generating biased Green Index scores. To mitigate this, model training should incorporate stratified sampling across market capitalizations and industry classifications, with regular bias audits by independent reviewers (Ong et al., 2025).

Second, corporate data security is paramount. The platform processes publicly available sustainability report data; however, any future integration of real-time operational data via corporate APIs must comply with Indonesia's Personal Data Protection Law (UU No. 27/2022) and relevant OJK data governance protocols. A clear data minimization principle must be adopted to ensure that only ESG-relevant data is extracted and retained.

Third, algorithmic transparency must be maintained through the implementation of Explainable AI (XAI) mechanisms, such as SHAP (SHapley Additive exPlanations) values, to ensure that all Green Index scores are accompanied by interpretable rationales accessible to non-technical stakeholders including retail investors, regulators, and the general public. This is particularly critical to prevent the Green Index from becoming a new "black box" that replicates the opacity it was designed to overcome (Sharma & Kumar Pandey, 2025; Mustafa et al., 2025).

4. Conclusions

In conclusion, the Green Balance innovation emerges as a transformative strategic solution that redefines the interaction between stakeholders and sustainability data. By harnessing the power of Artificial Intelligence, specifically through NLP and Machine Learning, this platform successfully automates the extraction of static data, standardizes diverse ESG metrics across industrial sectors, and quantifies corporate performance into an objective Green Index score. The systematic mapping of this innovation integrated through the SDLC framework, PESTEL analysis, and the Penta Helix collaboration model—demonstrates that Green Balance is not merely an idealistic concept, but a realistic and futuristic infrastructure ready for implementation in the Indonesian capital market.

This study fundamentally addresses the critical issue of information asymmetry by empowering stakeholders with a tool to validate corporate claims and make data-driven decisions based on measurable evidence. From a theoretical perspective, this research extends the application of Stakeholder Theory to the digital accounting realm, demonstrating that technological intervention is crucial for fulfilling information rights in complex reporting environments. Practically, the implementation of Green Balance is expected to significantly mitigate greenwashing practices, thereby directing capital flows toward a more transparent, accountable, and sustainable economic ecosystem. Ultimately, this innovation serves as a vital catalyst in accelerating Indonesia's transition toward its 2060 Net Zero Emission targets, bridging the gap between regulatory mandates and technological execution.

Acknowledgement

The author would like to thank all parties who contributed to the completion of this research

Author Contribution
Introduction, T.S.F; Methodology, L.L.A; Results and Discussion, T.S.F.; Design, N.W

Funding
This research received no external funding.

Ethical Review Board Statement
Not available.

Informed Consent Statement
Not available.

Data Availability Statement
Not available.

Conflicts of Interest
The authors declare no conflict of interest.

References
Adeyemo, A. (2025). *AI-powered sustainability reporting for carbon disclosure and ESG compliance*.

Aruwaji, M. A., & Swanepoel, M. J. (2025). The impact of AI-integrated ESG reporting on firm valuation in emerging markets: A multimodal analytical approach. *Journal of Risk and Financial Management, 18*(12), 675. https://doi.org/10.3390/jrfm18120675

Asia Investor Group on Climate Change. (2025). *State of investor climate transition 2025* (pp. 1–14). https://aigcc.net/wp-content/uploads/2025/07/Korea_Climate-Transition-Report-market-summary-English.pdf

Bais, B., Nassimbeni, G., & Orzes, G. (2024). Global Reporting Initiative: Literature review and research directions. *Journal of Cleaner Production, 471*, 143428. https://doi.org/10.1016/j.jclepro.2024.143428

Berniak-Woźny, J. (2025). The role of AI in ESG and sustainability reporting: A bibliometric study. *Economics and Innovation Studies, 3*(3), 1–15. https://doi.org/10.34659/eis.2025.94.3.1167

Bhattacharya, C. B., & Zaman, M. (2023). The what, why and how of ESG dashboards. *NIM Marketing Intelligence Review, 15*(1), 32–39. https://doi.org/10.2478/nimmir-2023-0005

Brickson, S. (2000). The impact of identity orientation on individual and organizational outcomes in demographically diverse settings. *Academy of Management Review, 25*(1), 82–101. https://doi.org/10.2307/259264

Chen, L., Dai, T., Zhang, C., & Zhang, Z. (2025). Digital government and corporate ESG performance. *International Review of Financial Analysis, 105*, 104379. https://doi.org/10.1016/j.irfa.2025.104379

Dewi, F. S., & Dewayanto, T. (2024). Peran big data analytics, machine learning, dan artificial intelligence dalam pendeteksian financial fraud: A systematic literature review. *Diponegoro Journal of Accounting, 13*(3), 1–15. http://ejournal-s1.undip.ac.id/index.php/accounting

Dewi, S. (2017, January). *Agenda kebijakan*. Blogspot. https://silvanadewi09.blogspot.com/2017/01/agenda-kebijakan.html

Dicoding Indonesia. (2021). *Mengenal model SDLC: Waterfall*. Dicoding Blog. https://www.dicoding.com/blog/

Ekaristi, C. Y. D., Utomo, D. C., & Rohman, A. (2025). Big data and AI in ESG performance measurement: A bibliometric analysis. *Edelweiss Applied Science and Technology, 9*(5), 2732–2749. https://doi.org/10.55214/25768484.v9i5.7587

Fildisi, B., Vakaj, E., Dridi, A., Imran, A. S., & Azad, R. M. A. (2025). Integrating AI-driven analytics for enhanced ESG mapping: Aligning local and global perspectives. *Sustainable Futures, 10*, 101231. https://doi.org/10.1016/j.sftr.2025.101231

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman Publishing.

Gao, S. S., & Zhang, J. J. (2006). Stakeholder engagement, social auditing and corporate sustainability. *Business Process Management Journal, 12*(6), 722–740. https://doi.org/10.1108/14637150610710891

Gutterman, A. S. (2024). *Sustainability reporting frameworks, standards, instruments, and regulations: A guide for sustainable entrepreneurs*. Sustainable Entrepreneurship Project. https://dx.doi.org/10.2139/ssrn.3809288

Harnessing big data and AI to revolutionize sustainability accounting and integrated corporate financial reporting. (2025). *International Journal of Computer Applications Technology and Research*. https://doi.org/10.7753/IJCATR1406.1008

Katz, S., Gu, Y., & Jiang, L. (2024). *Information extraction from ESG reports using NLP: A ChatGPT comparison* [Working paper]. SSRN. https://doi.org/10.2139/SSRN.4836432

Kaur, A., & Lodhia, S. K. (2014). The state of disclosures on stakeholder engagement in sustainability reporting in Australian local councils. *Pacific Accounting Review, 26*(1–2), 54–74. https://doi.org/10.1108/PAR-07-2013-0064

Ljunggren, T. (2024). *AI-driven metric extraction of sustainability reports* (TRITA-EE Report No. 2024:0000). KTH Royal Institute of Technology.

Mohammadrezaei, M., Huq, A., & Marques, C. (2024). *Use of text mining and natural language processing techniques in analyzing sustainability reports: A systematic literature review and assessment*. ResearchGate. https://www.researchgate.net/publication/388830881

Mohd Yusof, A. F., & Widyasamratri, H. (2025). Environmental, social, and governance: A review of frameworks, metrics, and reporting for sustainable development. *Civil and Sustainable Urban Engineering, 5*(2), 102–116. https://doi.org/10.53623/csue.v5i2.809

Mustafa, F., Smolarski, J., & Elamer, A. A. (2025). The convergence of artificial intelligence and sustainability reporting: A systematic review of applications, challenges and future directions. *Business Strategy and the Environment, 34*(6), 9761–9784. https://doi.org/10.1002/bse.70090

Ong, K., Mao, R., Satapathy, R., Filho, R. S., Cambria, E., Sulaeman, J., & Mengaldo, G. (2025). Explainable natural language processing for corporate sustainability analysis. *Information Fusion, 115*, 102726. https://doi.org/10.1016/j.inffus.2024.102726

Otoritas Jasa Keuangan. (2017). *Peraturan Otoritas Jasa Keuangan Nomor 51/POJK.03/2017 tentang penerapan keuangan berkelanjutan bagi lembaga jasa keuangan, emiten, dan perusahaan publik*. Otoritas Jasa Keuangan.

Pande, S., & Mishra, A. (2025). Five decades of ESG reporting research: A synthesis and future research avenues. *Corporate Communications: An International Journal*, 1–24. https://doi.org/10.1108/CCIJ-01-2025-0020

Rahmania, S. K. (2025). Pengaruh sustainability report dalam meningkatkan nilai perusahaan: Studi literatur review. In *Seminar Nasional Pariwisata dan Kewirausahaan (SNPK)* (Vol. 4, pp. 869–873). https://doi.org/10.36441/snpk.vol4.2025.413

Ramdhan, D. (2025). The role of ESG disclosure in attracting sustainable investment in Indonesia's capital market. *RIGGS: Journal of Artificial Intelligence and Digital Business, 4*(3), 3384–3393. https://doi.org/10.31004/riggs.v4i3.2484

Salim Malik, A. R., Brahmbhatt, S., & Vinay, A. (2025). AI-enabled ESG reporting: Bridging sustainability and accounting. *IJSAT – International Journal on Science and Technology, 16*(3). https://doi.org/10.71097/IJSAT.V16.I3.7541

Sattar, M. U., Dattana, V., Hasan, R., Mahmood, S., Khan, H. W., & Hussain, S. (2025). Enhancing supply chain management: A comparative study of machine learning techniques with cost-accuracy and ESG-based evaluation for forecasting and risk mitigation. *Sustainability, 17*(13), 5772. https://doi.org/10.3390/su17135772

Setiatin, T. (2025). Sustainable accounting and environmental, social, and governance (ESG) reporting: Challenges and implementation in Indonesian companies. *International Journal of Science and Society, 7*(1), 756–770. https://doi.org/10.54783/ijsoc.v7i1.1453

Sharma, D., & Pandey, V. K. (2025). AI-driven environmental, social, and governance (ESG) metrics: Bridging the gap between ethical finance and ecological sustainability.

Sun, Z., Satapathy, R., Guo, D., Li, B., Liu, X., Zhang, Y., Tan, C. A., Filho, R. S., & Goh, R. S. M. (2024). Information extraction: Unstructured to structured for ESG reports. In *2024 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 487–495). IEEE. https://doi.org/10.1109/ICDMW65004.2024.00068

Survey: Internal audit use of artificial intelligence growing rapidly. (n.d.). *Internal Audit 360*. Retrieved December 26, 2025, from https://internalaudit360.com/survey-internal-audit-use-of-artificial-intelligence-growing-rapidly/

The Definition. (2023). *Political factors definition*. https://the-definition.com/term/political-factors

Tjahjadi, B., Soewarno, N., & Mustikaningtiyas, F. (2021). Good corporate governance and corporate sustainability performance in Indonesia: A triple bottom line approach. *Heliyon, 7*(3), e06453. https://doi.org/10.1016/j.heliyon.2021.e06453

Zadeh, M. E., Kambar, N., & Zadeh, E. (n.d.). *Harnessing NLP and large language models for pattern discovery and information extraction in electronic health records*. https://doi.org/10.34917/37714603

Zou, Y., Shi, M., Chen, Z., Deng, Z., Lei, Z., Zeng, Z., Yang, S., Tong, H., Xiao, L., & Zhou, W. (2025). ESGReveal: An LLM-based approach for extracting structured data from ESG reports. *Journal of Cleaner Production, 489*, 144572. https://doi.org/10.1016/j.jclepro.2024.144572

Biographies of Authors

Tiara Saharani Fatimah, Accounting Study Program, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia.
- Email: stiaraf1456@students.unnes.ac.id
- ORCID: N/A
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A

Leila Luthfia Ahnaf, Accounting Study Program, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia.
- Email: fialee24@students.unnes.ac.id
- ORCID: N/A
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A

Nur Wisawalisma, Accounting Study Program, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia.
- Email: nurwisawalisma@gmail.com
- ORCID: N/A
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A