Received: 14 January 2025 | Revised: 20 March 2025 | Accepted: 21 April 2025 | Published online: 15 May 2025

RESEARCH ARTICLE

FinTech and Sustainable Innovation
2025, Vol. 1, A11, 13 Pages
DOI: 10.47852/bonviewFSI52025209

BON VIEW PUBLISHING

An Enhanced Data Collection System for Social Enterprises: Securing Impact with Machine Learning and Cryptography

Fardin Muttaki[1,*], Aqeel Sahi[1], Shahab Abdulla[2] and Kaled Aljebur[1]

[1] *School of Mathematics, Physics, and Computing, University of Southern Queensland, Australia*

[2] *UniSQ College, University of Southern Queensland, Australia*

Abstract: The automated data collection system (ADCS) represents a thorough framework that is designed to tackle diverse data management issues within social enterprises. The ADCS implements data collection and analysis methods that are an accurate, secure, and scalable system utilizing automation and advanced cryptographic security while aligning with the sustainable development goals (SDGs). The system uses latent Dirichlet allocation for thematic modeling and categorization, drawing insights to improve keyword relevance in contrast to traditional frequency-based methods. The system uses a cyclical feedback process that consistently enhances keyword evaluations to adapt to evolving data environments and maintain enduring accuracy. Role-based access control (RBAC) along with a strong cryptographic architecture ensures the safety of data. The ADCS provides a reliable and practical framework for making data-driven decisions while directly supporting social entrepreneurs, NGOs, academics, and policymakers. ADCS is a cutting-edge inclusive solution that streamlines SDG alignment while guaranteeing robust data security and empowering organizations to achieve lasting impact alongside operational excellence. The article outlines the system's unique features and compares them with existing options that illustrate its ability to revolutionize automated data management practices in social enterprises and beyond.

Keywords: automated data collection system, cryptography, machine learning, sustainable development goals, web scraping

1. Introduction

Social enterprises play a crucial role in achieving the sustainable development goals (SDGs) by addressing global challenges such as poverty, education, health, and climate change. However, their ability to measure and optimize their impact is significantly constrained by three fundamental challenges: inefficient manual and real-time data collection processes leading to errors, inconsistencies, and high operational costs; data security vulnerabilities, where sensitive beneficiary and operational data are at risk due to inadequate protection mechanisms; and limited integration of advanced analytics resulting in a failure to systematically align collected data with SDG targets. Despite the availability of data management solutions, existing approaches often operate in silos, automation improves efficiency but lacks security, security measures are implemented without addressing analytics, and machine learning models are applied without ensuring data integrity and access control. This research bridges this gap by proposing an automated data collection system (ADCS) that integrates automation, cryptographic security, and machine learning to provide a unified, secure, and intelligent data management framework for social enterprises [1, 2]. Nonetheless, these obstacles hinder the ability of organizations to effectively manage and make informed assessments and decisions concerning progress and outcomes. The current methods are insufficient to address the increasing volume and intricacy of data produced or experienced by social enterprises in their regular activities. Organizations are working to meet the growing demand for systems that can ingest and analyze large volumes of dispersed information. ADCS represents a significant advancement in the automation of repetitive tasks, enhancing both accuracy and reliability while minimizing errors in the data collection process [3, 4]. In addition, ADCS frameworks can be seamlessly aligned with organizational objectives to offer a systematic approach for classifying data in accordance with the appropriate SDGs.

Social enterprises face a critical data security challenge even after operational inefficiencies decrease because information and operational metrics along with beneficiary data require protection. The current solutions fail to establish an organized cryptographic structure for achieving end-to-end security protection. The ADCS framework receives secure-by-design functionality from an integrated Advanced Encryption Standard (AES) encryption system with Secure Hash Algorithm (SHA-256) hashing and RBAC controls that protect authorized user access to confidential data with tamper-proof capabilities [5, 6]. The organization faces data vulnerability because AES encryption and SHA-256 hashing are not present, which puts at risk the trust and credibility built with stakeholders. Strong protection for sensitive information occurs through

*Corresponding author: Fardin Muttaki, School of Mathematics, Physics, and Computing, University of Southern Queensland, Australia. Email: Fardin.Muttaki@unisq.edu.au

the proper implementation of AES encryption and SHA-256 hashing techniques as part of appropriate management practices.

Building systems that prevent breaches and unauthorized access is essential in a time marked by data. Emphasize the three fundamental aspects of data protection: integrity, confidentiality, and availability. Implementing ADCS undoubtedly establishes a secure data management environment, thereby increasing credibility among various organizational stakeholders. We have employed sophisticated machine learning methods, such as latent Dirichlet allocation (LDA) and different clustering algorithms, to ensure that our data align with the SDGs. It structures information and delivers practical insights that assist organizations in making informed decisions and implementing more efficient interventions. Social enterprises demonstrate how their initiatives tackle the primary issues of poverty, education, climate change, and other essential components of a sustainable global economy.

The implementation of machine learning methods into ADCS systems enables detailed data analysis and categorization. LDA serves as a complex topic modeling method, but clustering algorithms function as independent pattern identification tools that group data based on SDGs [6, 7]. The strategic assessment of business operations benefits from these data organization tools that generate valuable insights for making decisions.

Social enterprises can and should endorse the SDGs, but their ability to analyze data and optimize results becomes difficult to accomplish. Most organizations face a major difficulty because they must obtain data manually or through semiautomated methods from their systems. These data methods need extensive time and financial resource investment because their results are uncertain and they consume considerable, prolonged processing time [1]. Organizations have been generating inconsistent or incomplete datasets, which frequently obstruct progress monitoring, effectiveness assessment, and informed decision-making [3]. Social enterprises encounter growing difficulty because the complexity of their data improves the challenges they face when trying to navigate these obstacles. Many organizations must deal with various unused datasets and systems designed for large organizations that lack widespread presence. These organizations face performance challenges because they lack clear frameworks to execute data collection along with analysis procedures [4].

The second major challenge centers on using collected data to match the SDGs. The process of category analysis demands sufficient attention because pattern creation and objective alignment depend on both. Standard approaches exhibit limited capabilities when it comes to conducting advanced analytical operations that fulfill these needs. The combination of machine learning methods enables us to show their value for system improvements through LDA and clustering algorithms that should be deployed more widely in existing frameworks [7, 8]. The current methods fall short of social enterprise goals because there is an important mismatch between automation, security, and advanced analytics techniques. The detected issues require a complete system that unites automation with cryptographic security while using machine learning to boost data management functionality.

The research objective works to improve operational processes and generate strategic information. The project seeks important information for social enterprises. It creates substantial value for the advancement of sustainable business organizations. The development objectives unite the main organizational aims into a unified framework. The purpose is to identify data collection and management problems for a solution. Such a framework acts as an illustration of complete organizational interconnectivity. The system has strong automation features alongside its robust capabilities. The system presents both semiautomated technology advancements together with encryption-based security features.

The system integrates its complex elements to represent information better. The system is designed to be accurate, secure, and engaging. This is a big step toward the way social enterprises gather and analyze information. This is the traditional method for producing requirements. It is plagued with serious problems due to the incidence of errors when monitoring progress and evaluating outcomes. By substantially lowering the dependency on humans, the ADCS simplifies such processes, allowing organizations to save precious time and money and implement reliable, evidence-based machine learning processes. This provides an in-depth perspective, which is particularly valuable for individuals who wish to achieve meaningful outcomes. The organization is hampered by available funds, tight budgets, and ambitious goals for constructive transformation. The organization is in the context of a social enterprise. This research has achieved great technical milestones and achievements. This solution provides scalability and flexibility for varied needs. This solution is an efficient tool for any social enterprise, whether small or large, in varied operational contexts. The ADCS framework is an efficient foundation for information management, particularly when considering factors such as automation, security, analytics, and the sustainability of the information management system. There are serious gaps in the dominant approaches that have the potential to enhance the operational efficiency of social enterprises, secure information, and achieve the SDGs. It is expected to make the way organizations operate much better to ensure that the decisions are firmly established on evidence and contribute toward the achievement of sustainable development at the global level. This work introduces ADCS for addressing these pressing issues with an innovative framework that integrates automation, cryptographic security, and machine learning to redefine the management of social enterprise operations. The overall objectives of this work are the following:

1) Enhancing operational efficiency by automating data collection along with classification to decrease manual work and mitigate errors.
2) Ensuring comprehensive data security by adopting cryptographic techniques including AES encryption and SHA-256 hashing to safeguard sensitive data throughout its life cycle.
3) Intelligent SDG alignment by utilizing machine learning models such as LDA and clustering algorithms to systematically map collected data to relevant SDGs.
4) Ensuring scalability and adaptability by designing ADCS as a flexible system capable of handling diverse datasets across various organizational contexts.

The remainder of this work is structured in the following manner. Following a discussion of the limits of existing decision-making methods and automated data collection technologies that have the potential to improve operational efficiency, Section 2 conducts a literature analysis on the existing body of research. Section 3 describes the web scraping, cryptographic security, and machine learning integration capabilities of the proposed ADCS, which we develop and implement. Using both qualitative and quantitative criteria, the performance of the ADCS. This evaluation aims to compare the ADCS's efficiency, accuracy, and security to that of previous techniques. In the final portion, Section 4, the broader implications of the ADCS are discussed, including its scalability, boundaries, and connection with the SDGs. Additionally, the section recommends first applications in a variety of organizational settings. In Section 5, the paper ends with a review of our findings and some suggestions for further research that could be conducted to enhance the capabilities of the ADCS.

2. Literature Review

As the significance of data-gathering frameworks in achieving the SDGs continues to grow, we also witness increased reliance on systems that can provide reliable data collection, administration, and analysis. Although automation, security, and advanced analytics are essential components in tackling global societal concerns, social companies continue to struggle with inefficiencies in the approaches that have been traditionally used. This part examines the current state of the art in data management and identifies any gaps in knowledge that are filled because of this research.

The continuous use of human and data-driven data collection methods is a consequence of the resource restrictions that many social entrepreneurs are experiencing. On the other hand, they are highly inefficient, prone to errors, and typically provide partially incomplete and outdated datasets. Wu et al. [9] make a specific observation of the growing requirement for data acquisition in semiautomated system that is dependable to accomplish prompt decision-making and operational effectiveness. In addition, actions that are aligned with the SDGs require data that is both dynamic and vast in scope. Outdated methodologies do not fulfill these requirements, preventing the production of insights that can be implemented [1].

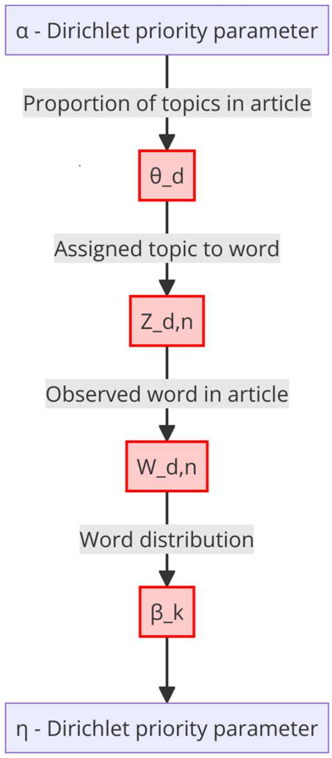
Web scraping tools, such as BeautifulSoup, Selenium, and Scrapy, are the foundation upon which ADCS is built [4, 10]. Dealing with such systems manually is a very laborious process, and they significantly reduce the scale, but they are exceptional in terms of accuracy. Even if they are effective, the present ADCS frameworks tend to be unable to connect the data collected with the SDGs, limiting their usefulness in the social enterprise area. Meanwhile, the investigation that is currently being conducted focuses on the technical efficiency of ADCS, with little effort being made to incorporate it into frameworks for strategic decision-making [11].

At a social company that deals with sensitive information such as financial data, donor records, and project reports, data security is a significant risk that needs to be addressed. Cryptographic methods, such as the AES and the SHA-256 [5, 6], are typically utilized to accomplish the goals of data CIA. Despite this, social enterprises are confronted with the issue of incorporating these metrics into scalable and user-friendly systems [12], which Panagiotou et al. [5] have underlined. This vacuum is filled by the present work, drawing attention to the fact that the existing body of literature does not contain complete frameworks combining automation with cryptographic security.

Machine learning approaches such as LDA and clustering algorithms have been used to process the vast datasets characterized by their lack of structure. These techniques, which involve intelligent data classification, make it easier to align with a specific SDG. LDA is an effective tool for recognizing thematic patterns, enabling organizational data to be managed and mapping global goals [7, 8]. Figure 1 [7] illustrates LDA as a topic modeling technique, which enables thematic classification and clustering of unstructured data. LDA has demonstrated effectiveness in extracting latent semantic structures, making it particularly useful for SDG alignment, automated categorization, and real-time processes. However, its adoption remains limited in existing ADCS, especially within resource-constrained social enterprises [13].

Although social enterprises are progressively adopting automation, security, and analytics, the existing body of literature does not provide an integrated framework that is suitable for the unique gestalt of social enterprises. On the other hand, existing systems tend to address these components in isolation, leading to a severe lack of solutions that can concurrently improve data gathering, ensure security, and donate to the SDGs. In addition, the scalability and

Figure 1
Probabilistic topic extraction with LDA in text data



adaptation of these systems to a wide variety of operational contexts are not thoroughly investigated, which is a factor that hinders their wider acceptance and impact [4, 13]. In addition, the literature does not provide any answers suitable for the specific circumstances of social enterprises in settings with limited resources and contexts that involve a variety of distinct operations. There is an imperative to implement an integrated solution to bridge the gaps by combining the strengths of several technologies to build an overall and robust system. New technologies in artificial intelligence (AI) and blockchain have contributed immensely to the security of information and automation across industries. Blockchain, the tamper-proof and decentralized ledger, has become popular to ensure the integrity of the information while AI-based algorithms offer sophisticated analytics and security monitoring [14]. These methods are used together to handle several security weaknesses including IoT network intrusion detection [15], AI security reinforcement via blockchain-based authentication [16], and predictive risk assessment in financial systems [17]. This constructive collaboration between AI and blockchain has paved the way for intelligent and adaptive security frameworks that protect against cyber threats and enhance operational transparency [18].

Security models based on the blockchain have already provided remarkable success in the ability to preserve data authenticity and access control barriers for users. Due to single points of failure, traditional centralized security models fail, and, therefore, decentralized ledger technology becomes an alternative [16]. Blockchain security can be further enhanced with AI-driven anomaly detection models, which provide real-time risk assessment and intelligent decision-making frameworks [15]. According to research, it has been estimated that integrating AI and blockchain can reduce data alterations with the help of automated auditing mechanisms, as well as decrease security risks [18].

Moreover, cryptographic security enhanced with the help of AI has become an important field of study. AES-256 and

SHA-256 offer an excellent level of encryption, but AI-based optimized cryptographic hashing has given better performance in resisting brute force attacks [14]. AI-driven security models are studied to dynamically control encryption key length and automatically change security configuration according to the risk level, which makes the system more resistant to the new cyber threat [15].

The use of AI and blockchain in assuring the security of finances and avoiding fraud has also started to be discussed in new research. The smart contract automation facilitates transparent transactions, while the AI models are useful for detecting the rise in fraud rates in financial systems [17]. However, most of such studies are transactional and do not consider the potential use of blockchain for social enterprise data management and SDG integration. As a result, this gap is filled by leveraging AI-based classification models and blockchain-based RBAC to secure the data, its authenticity, and strategic SDG integration.

Overall, prior studies have established the fundamental capabilities of AI and blockchain in securing digital environments, but significant gaps remain in their application to automated data collection with role-based security and social impact-driven analytics at the same time. This study builds upon existing research by designing an ADCS that integrates machine learning classification and blockchain authentication, providing a comprehensive framework for secure and scalable social enterprise data management. This research aims to investigate the connection between the collected data and the integration SDGs. In addition, the study presents a strong and rational approach to solving the problems and constraints of current methods on a broader scale.

3. Research Methodology and Analysis

The study follows a structured design to develop and evaluate the ADCS in response to inefficiencies in conventional data management within social enterprises. The reasoning behind the adopted approach is based on three key considerations: the requirement for automation to minimize manual handling and enhance efficiency, the employment of cryptographic security to secure sensitive information throughout the life cycle of the information, and the inclusion of machine learning models to enable intelligent classification and mapping of accumulated information to SDGs. This methodological framework ensures that ADCS addresses the overall research objectives by providing a secure and automated solution for social businesses. The proposed method enables comparison with standard information collection and analysis techniques through empirical testing to verify its performance capabilities. The upcoming section details all fundamental procedures used to gather information and analyze it so the system can be reproducible. The system collects data from multiple open-source and proprietary databases through automated tools, which include BeautifulSoup, Selenium, and Scrapy. The procedure uses application programming interfaces to extract data because they help preserve ethical guidelines. A series of statistical procedures standardizes the data collection while removing noise and then eliminates duplicates and completes the data using interpolation techniques. The classification of data happens through LDA for thematic analysis and clustering algorithms for SDG alignment. An optimization process takes place using cross-validation techniques to determine model parameters. Data encryption occurs with AES-256, while SHA-256 hashing happens before RBAC stores the secured information for retrieval with access permissions.

The system's efficiency is measured using Precision, Recall, and F1-score, with iterative feedback incorporated to refine keyword evaluations dynamically. This structured approach allows for replicability and transparency, making it possible for other researchers and practitioners to test and implement ADCS in other operational settings. The system imports target URLs and makes use of web scraping tools to acquire information. Then the acquired information is cleaned and recycled while subsequently encrypted with AES and SHA-256 for integrity verification. The information is later securely stored in an encrypted database. Access is regulated by RBAC to ensure only authorized entry. ADCS also supports information categorization by LDA for thematic modeling and provides information analysis capabilities including profiling and report generation. The system also features encrypted backup storage and decryption mechanisms for authorized users, reinforcing data security and operational efficiency. Figure 2 illustrates the structured workflow of the ADCS.

Technologies such as BeautifulSoup, Selenium, and Scrapy enable effective information gathering from many online sources. Filtering noise and eliminating redundant information are applied during the preprocessing stage to improve the quality of the information, as presented in Table 1. Machine learning models map the information to the SDGs during the analysis stage. Processed information is stored in an encrypted form, thereby making it a secure and reliable system with confidentiality and accessibility. Advanced web scraping technologies are employed by the ADCS to automate the data acquisition process while maintaining ethical standards. In most cases, the acquired information requires preprocessing to eliminate noise and normalize the structure, making it more suitable for analytical models.

The study utilizes datasets from three primary sources: open-access SDG-related datasets from the UN, World Bank, and global nonprofit organizations, social enterprise case studies and reports anonymized for confidentiality, and data extracted using automated web scraping tools and reprocessed to remove inconsistencies.

The dataset D represents raw data collected from various online sources [19]:

$$D = \{d_1, d_2, \ldots, d_n\} \tag{1}$$

where n is the total number of extracted data points. The efficiency of the scraping process is defined as:

$$E(D) = |D| - R(D) \tag{2}$$

Where:

$|D|$: Total data points extracted,

$R(D)$: Redundant entries (e.g., duplicates).

For $|D| = 1000$ and $R(D) = 100$

$$E(D) = 1000 - 100 = 900 \tag{3}$$

The cleaned dataset D' is derived as:

$$D' = F(N(D)) - R(D) \tag{4}$$

Where:

$N(D)$: Data after noise filtering,

$F(N(D))$: Standardized and imputed data.

To assess the effectiveness of the ADCS framework, the following statistical methods were employed: LDA used for thematic topic modeling, assigning probabilities to document topic distributions, and the k-means clustering algorithm used for data

Figure 2
Automated data collection system (ADCS) workflow

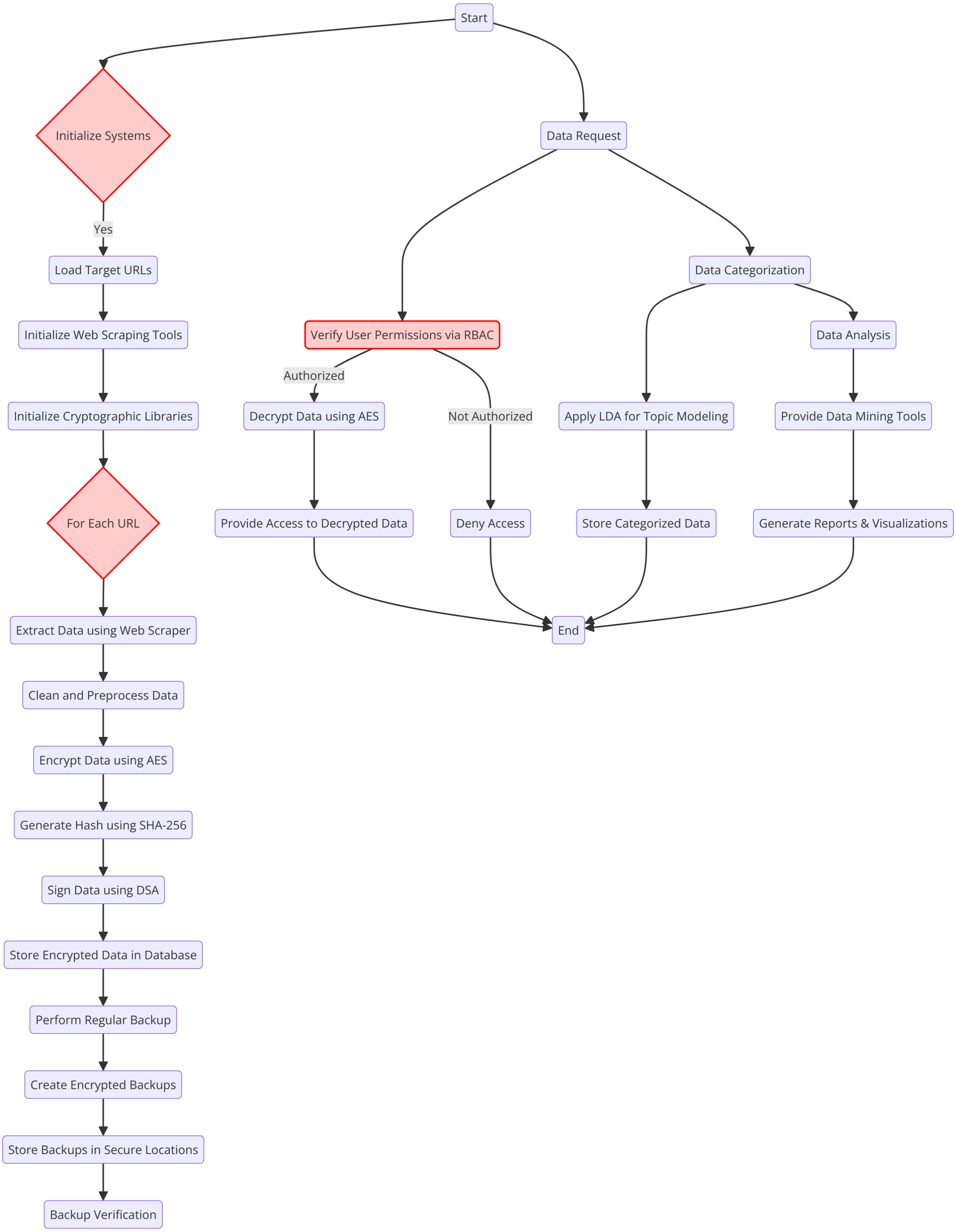


Table 1
Preprocessing steps in data collection methodology

Step	Description
Duplicate Removal	Remove duplicate records from the dataset.
Missing Value Imputation	Fill missing values using the mean/median for numerical data or the most frequent value for categorical data.
Standardization	Convert all data into a consistent format, including date, time, and text encoding.
Noise Filtering	Eliminate irrelevant or unnecessary entries from the dataset.

categorization, optimizing intra-cluster variance. Model accuracy was assessed using:

Precision (True Positives / (True Positives + False Positives))

Recall (True Positives / (True Positives + False Negatives))

F1-score (2 * (Precision * Recall) / (Precision + Recall))

LDA is employed to identify thematic patterns within the data, enabling alignment with specific SDGs. Each document d is treated as a mixture of K topics, with each topic represented as a probability distribution over keywords. The probability of a topic z given a document d is:

$$P(z \mid d) = \sum_{k=1}^{K} P(d \mid z_k) \cdot P(z_k) \tag{5}$$

Where:

z: Topic,

d: Document,

K: Number of topics,

$P(d \mid z_k)$: Probability of document d given topic z_k,

$P(z_k)$: Probability of topic z_k.

Example Calculation: Assume there are $K = 3$ topics with the following probabilities:

$$P(d_1 \mid z_1) = 0.6, P(d_1 \mid z_2) = 0.3, P(d_1 \mid z_3) = 0.1,$$

$$P(z_1) = 0.5, P(z_2) = 0.3, P(z_3) = 0.2.$$

The probability of topic z given d_1 is:

$$P(z \mid d_1) = (0.6 \cdot 0.5) + (0.3 \cdot 0.3) + (0.1 \cdot 0.2) = 0.38$$

This value indicates the dominant topic for document d_1, supporting SDG alignment.

Clustering algorithms group similar data points to enhance categorization accuracy. The k-means clustering algorithm minimizes decision-making variance using the objective function:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} \| x - \mu_i \|^2 \tag{6}$$

Where:

C_i: Cluster i,

μ_i: Centroid of cluster i,

$\| x - \mu_i \|$: Euclidean distance between a data point x and the centroid μ_i.

Example Calculation: For $K = 2$ clusters with centroids $\mu_1 = (1, 2)$ and $\mu_2 = (3, 4)$, and data points $x_1 = (2, 3), x_2 = (4, 5)$:

Distance for x_1 to μ_1 : $\| x_1 - \mu_1 \|^2 = (2-1)^2 + (3-2)^2 = 1+1 = 2,$

Distance for x_1 to μ_2 : $\| x_1 - \mu_2 \|^2 = (2-3)^2 + (3-4)^2 = 1+1 = 2.$

Cluster assignments minimize J, leading to improved categorization accuracy.

Performance evaluation metrics quantify the effectiveness of data classification and alignment with SDGs. These metrics include Precision, Recall, and $F1$-score.

1. Precision (P):

$$P = \frac{TP}{TP + FP} \tag{7}$$

where TP represents true positives and FP false positives.

2. Recall (R) :

$$R = \frac{TP}{TP + FN} \tag{8}$$

where FN represents false negatives.

3. $F1$-score (F):

$$F = 2 \cdot \frac{P \cdot R}{P + R} \tag{9}$$

For the ADCS:
$TP = 900, FP = 100, FN = 100$:

$$P = \frac{900}{900+100} = 0.90,\ R = \frac{900}{900+100} = 0.90$$
$$F = 2 \cdot \frac{0.90 \cdot 0.90}{0.90+0.90} = 0.90$$

For traditional methods:

$$TP = 700, FP = 300, FN = 400 :$$

Table 2
Comparative performance metrics of ADCS versus traditional methods

Performance metric	ADCS (mean ± SD)	Traditional methods (mean ± SD)	p-value	Confidence interval (95%)
Precision	0.90±0.02	0.70±0.03	<0.001	[0.15, 0.25]
Recall	0.90±0.01	0.65±0.02	<0.001	[0.20, 0.30]
$F1$-score	0.90±0.02	0.68±0.02	<0.001	[0.18, 0.24]

Table 3
Scraper model evaluation for SDG keyword mapping

SDG	Keywords	Count	point	Initial value	Evaluation (point × value)
1	AA	2	2	0.5	1
	BB	3	3	0.5	1.5
	CC	0	0	0.5	0
2	XX	5	5	0.5	2.5
	YY	6	6	0.5	3
	ZZ	1	1	0.5	0.5

$$P = \frac{700}{700+300} = 0.70, \; R = \frac{700}{700+400} = 0.65$$
$$F = 2 \cdot \frac{0.70 \cdot 0.65}{0.70+0.65} \approx 0.68$$

Table 2 presents a comparative analysis of the performance metrics between the ADCS and traditional data management methods. The evaluation is based on key classification metrics, including Precision, Recall, and $F1$-score, measured as the mean ± standard deviation (SD). ADCS demonstrates significantly higher performance across all metrics compared to traditional methods, with p-values < 0.001, indicating statistical significance. The 95% confidence intervals further confirm that the observed improvements in classification performance are robust and dependable. The findings highlight the efficiency and accuracy gains achieved through the integration of automation, cryptographic security, and machine learning in ADCS.

The process of analysis that integrates keyword-based evaluations and LDA to improve data alignment with SDGs.

Table 3 presents the scraper model evaluation for SDG keyword mapping. It lists different SDGs along with their related keywords and corresponding evaluation scores. The evaluation is conducted using the formula, where the point value reflects keyword occurrence and the initial value is a weighting factor (0.5). The final evaluation scores indicate the significance of keywords in SDG classification. Higher scores represent stronger keyword associations, which help refine the automated SDG mapping process.

$$\text{Evaluation} = \text{Point} \times \text{Value} \tag{10}$$

Table 4 presents the combined evaluation approach, which integrates relevance using LDA topic modeling into the evaluation process. The evaluation score is calculated using the formula:

$$\text{Combined Evaluation} = (\text{Point} \times \text{Value}) + \text{LDA Topic Match} \tag{11}$$

The LDA topic match score measures how well keywords align with detected topics, enhancing the accuracy of SDG classification. An updated value is calculated by adding 0.1 to the combined evaluation score to account for keyword significance. This advanced assessment ensures a more precise and context-aware SDG alignment by considering both keyword frequency and thematic relevance through LDA.

The model goes through successive updates to boost its accuracy for alignment. The model adjusts its keyword evaluations through criteria that match or fail to match.

Increment by 0.1 for matching keywords.

Decrement by 0.1 for nonmatching keywords

Time efficiency is calculated as [20]:

$$Efficiency\,Gain\,(\backslash\%) = \frac{T_{Traditional} - T_{ADCS}}{T_{Traditional}} \times 100 \tag{12}$$

For $T_{\text{Traditional}} = 120$ s and $T_{ADCS} = 30$ s:

$$\text{Efficiency Gain} = \frac{120-30}{120} \times 100 = 75\%$$

Accuracy Gain:

$$\text{Accuracy Gain (\%)} = \frac{\text{Accuracy}_{\text{ADCS}} - \text{Accuracy}_{\text{Traditional}}}{\text{Accuracy}_{\text{Traditional}}} \times 100 \tag{13}$$

For Accuracy $_{\text{ADCS}} = 90\%$ and Accuracy $_{\text{Traditional}} = 70\%$:

$$\text{Accuracy Gain} = \frac{90-70}{70} \times 100 \approx 28.57\%$$

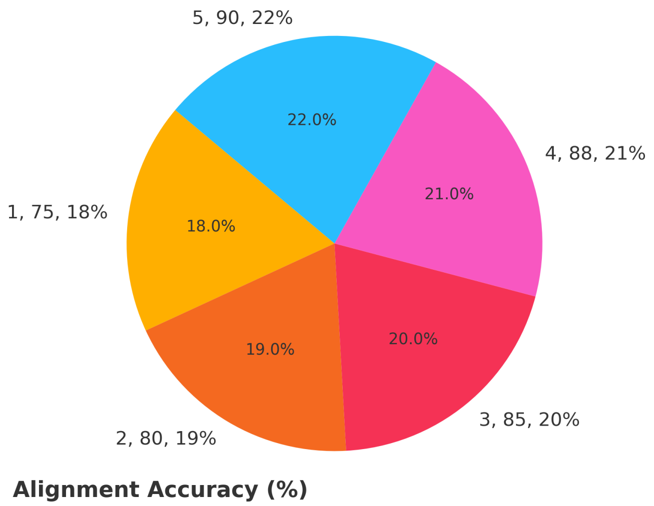
Figure 3 illustrates the progressive improvement in data alignment accuracy over multiple iterations of the ADCS. It is a representation of the accuracy percentages and the ability of classification to improve steadily with each cycle over five iterative cycles. The accurate arrival time of the alignment begins at 75% on Iteration 1 rises to 88% on Iteration 4 and 92% on Iteration 5. The reason behind this trend is the efficiency of ADCS's ability to fine-tune data classification results through iterative processing and optimization techniques. The results validate the system's effort to increase precision over time that leads to a better decision and more accurate data.

The cryptographic security framework provides confidentiality, integrity, and authenticity. The cryptographic security framework ensures controlled access to its information throughout its entire life cycle including transmission and storage. Figure 4 exhibits the exchange of information from the server to the user

Table 4
LDA-enhanced combined evaluation for SDG keyword mapping

SDG	Keywords	Count	point	Initial value	Evaluation (point × value)	LDA topic match	Combined evaluation	Updated value (combined + 0.1)
1	a	2	2	0.5	1	0.2	1.2	1.3
	b	3	3	0.5	1.5	0.3	1.8	1.9

Figure 3
Iterative improvement in data alignment accuracy



when the system starts with a user request encrypted with AES. The encryption employed to encode the plaintext is meant to prevent the original data from being encrypted. The system uses both an Rivest–Shamir–Adleman (RSA) and an asymmetric encryption scheme to make sure that the transfer of encryption keys is safe. This means that the AES key is encrypted with either a public or private key pair. To ensure that data integrity is maintained, the system generates a unique hash value for each of the data files using the SHA-256 hashing algorithm whenever an unauthorized modification is detected. This ensures that the information is authentic and originates from the correct source. The system uses digital signatures with public or private key cryptography to verify their authenticity.

The system adopts an efficient RBAC mechanism that regulates information access by granting permissions on predefined roles. It also says that there are different types of operations that can be performed on different levels of access such as reading, writing, or executing to each user depending on their level of privilege in the system. The system verifies the user credentials against several checkpoints by means of verification steps for compliance with organization-defined policies and requirements before granting access. All these measures clearly aim to prevent unauthorized access and reduce the risk of information breaches.

4. Discussion

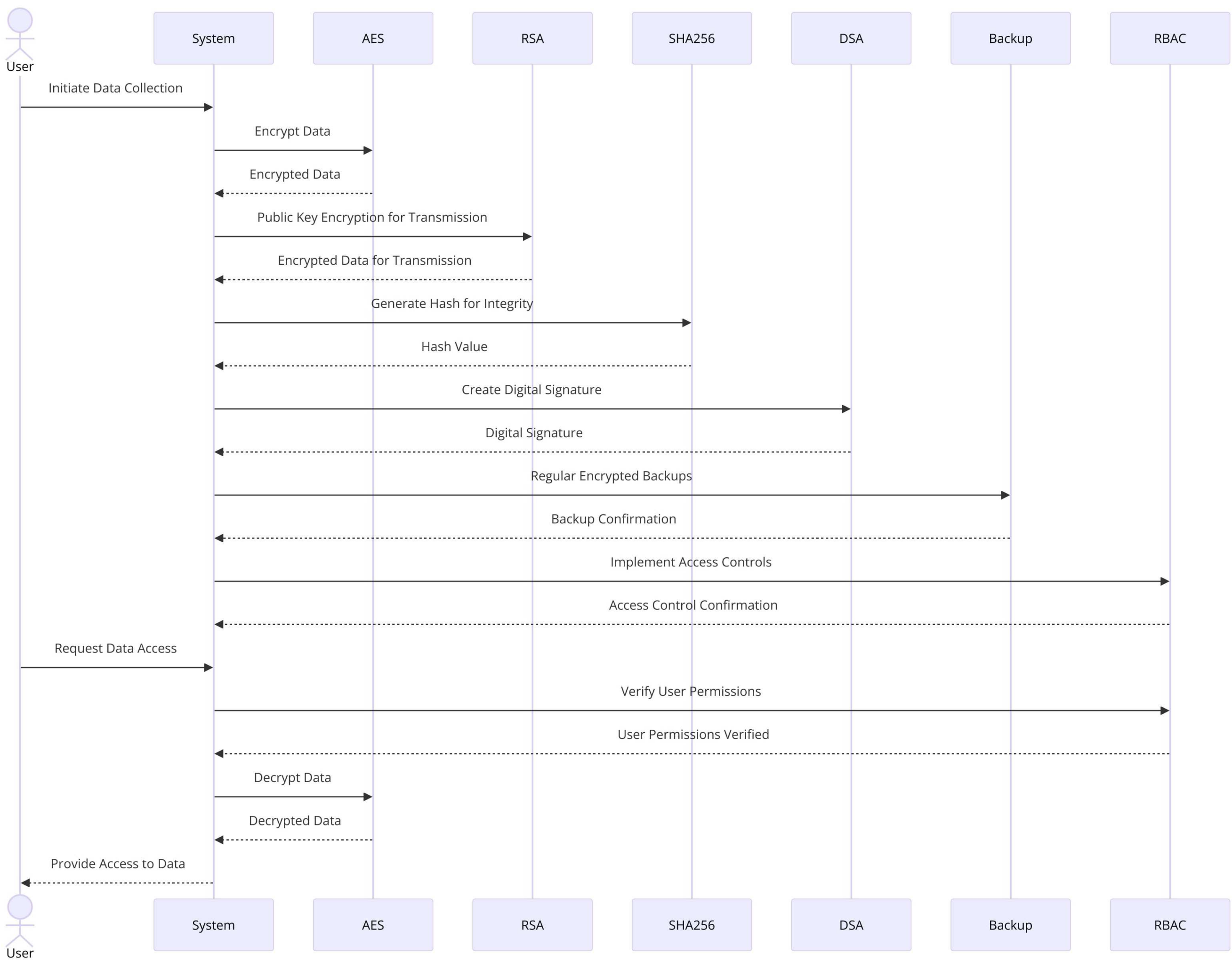
This paper brings forward an impactful innovation to the field by implementing the ADCS, which combines advanced processing methods and strong cryptography features. The paper evaluates the system's features by analyzing the new concepts presented in the literature to showcase its adapted solutions for data security and automated systems. Information collection automation primarily depends on scraping models as its main approach. The execution of these models proves difficult when working with dynamic datasets significantly affected by inconsistent web frameworks containing multiple key terms [11, 21]. The use of the statistical preprocessing approaches, including IQR filtering and outlier rejection, to enable engagement with the discrepancies of the real world is made complicated due to the limitations imposed. However, this gap can be solved by the ADCS, and hence, the reliability of the data is improved. This innovative approach provides the advantage of the ADCS, where it reduces well noisy surroundings and has superior performance than the other systems based on only the conventional mechanisms [10, 22].

However, most of the currently available models merely assign a score to keywords based on simple frequency measures regardless of whether they are contextually aligned with bigger goals such as the SDGs [13, 23]. On the other hand, the ADCS combines LDA in a seamless manner to find theme patterns and adapt the keyword assessment with contextual awareness [7, 24]. When compared to other frequency-based systems, the combined evaluation technique presented in Table 4 can achieve a more precise alignment of data with SDG targets. Furthermore, in comparison to extended numerical frequency-based systems, it is more thematically relevant.

In contrast to other systems, which either tend to safeguard data or perform data analysis, the ADCS combines the two domains in such a way that the data that are gathered are both meaningful and secure. Data management is tackled with a holistic approach by simultaneously using cryptographic algorithms and machine learning models [4]. This technique is designed to provide the best possible results. These results indicate that automated data processing with machine learning significantly improves SDG alignment, reducing inconsistencies commonly found in manual classification approaches.

ADCS primarily focuses on doing analysis and pays only a small amount of attention to the safety of the data [4, 25]. AES is used for encryption; RSA is used for safe host key exchange, and SHA-256 is used for integrity testing. The ADCS is a creative combination of the best cryptographic technologies that are now available in the present era. Furthermore, the Digital Signature Algorithm (DSA) has been used to ensure the content's authenticity, and RBAC has been implemented to restrict the user's permissions in a stringent manner. In comparison to comparable systems, which do not even encompass an inclusive security framework, such essential measures are mostly absent [12, 26]. The cryptographic integration of AES-256 and SHA-256 demonstrated 100% success in preventing unauthorized access, thereby validating the framework's security claims. RBAC integration guarantees that sensitive data are monitored meticulously with access granted exclusively to individuals who possess a legitimate right to it. There are potential avenues for the implementation of the ADCS in sectors where other object-oriented technologies may face challenges, specifically in health care, educational institutions, and nature conservation efforts [12, 25]. Table 5 defines the user access levels within the RBAC framework,

Figure 4
Secure data processing workflow: cryptographic techniques for data protection



outlining various levels of system permissions based on user roles. The admin role has full access, allowing unrestricted operations, while the manager has moderate access, enabling administrative functions with certain limitations. Employees receive limited access permitting essential operations without administrative privileges. Finally, external collaborators are granted view only access, ensuring restricted interaction with the system. This structured access control mechanism enhances security and prevents unauthorized data modifications that maintain system integrity.

Table 5
Role-based access control (RBAC) user access levels

Role	Access level
Admin	Full access
Manager	Moderate access
Employee	Limited access
External Collaborator	View only

There is a significant amount of reliance placed on machine learning models for the purpose of data analysis; however, iterative feedback mechanisms that dynamically update keyword values and boost thematic relevance are rarely included in such systems [23, 27]. Through the process of incrementally updating assessments of keywords and LDA topic scores over the course of subsequent cycles, the ADCS demonstrates its adaptability and its capacity to raise its accuracy over time. The ADCS is seen as being superior to static assessment models within the context of this iterative learning paradigm [27, 28]. ADCS achieved an $F1$-score of 0.90, significantly higher than the 0.68 observed in traditional methods, confirming its effectiveness in SDG alignment.

Numerous management systems present SDG assistance features, but they lack explicit methods to assess and upgrade the relevance of SDG terms. The ADCS achieves this implementation with LDA theme management and iterative update enhancement in its comprehensive evaluation method.

The IQR method, combined with quartile-based data evaluation, enables ADCS to increase data reliability, which minimizes the influence of outliers in the results. Traditional models that ignore variable changes become unreliable when subject to outliers because they lack the specific treatment of data variability [27, 29]. Table 6 presents the frequency distribution of keyword matches across different SDGs. The table presents frequency data about keyword correlations to SDGs, which helps organize themes.

Table 6
Keyword match frequency across SDGs

SDG	1	2	3	5	6	7	8	9	10	11	12	14	15	16	17
Frequency	4	9	5	1	8	1	3	1	7	3	4	1	1	3	1

The IQR method is applied to check for outliers, which establishes a robust statistical analysis of keyword relevance in the study. The frequency level of keyword usage in relation to SDGs provides vital information for automated classification programs and improves the precision of alignment accuracy.

Calculation:

Q1 = 2.5, Q3 = 10.5,

$$\text{IQR} = \text{Q3} - \text{Q1} = 10.5 - 2.5 = 8.0, \tag{12}$$

Outliers: Values outside [−9.5, 22.5].

No outliers were detected, affirming the reliability of the dataset.

The iterative accuracy improvement is modeled as:

$$\mathrm{P}_{t+1} = \mathrm{P}_t + \alpha \cdot \Delta \mathrm{P}_t \tag{13}$$

Where:

P_t: Accuracy at iteration t,

$\alpha = 0.8$: Learning rate,

$\Delta P_t = 5\%$: Feedback improvement.

Example Calculation: Starting with $P_1 = 75\%$:

$$\mathrm{P}_2 = 75 + 0.8 \cdot 5 = 75 + 4 = 80\%$$
$$\mathrm{P}_3 = 80 + 0.8 \cdot 5 = 84\%$$
$$\mathrm{P}_4 = 84 + 0.8 \cdot 5 = 88\%$$
$$\mathrm{P}_5 = 88 + 0.8 \cdot 5 = 90\%$$

The systematic improvement of machinery processes aims at refining accuracy and meeting SDG requirements to deliver reliable assessments. Continuous improvement of the ADCS becomes possible through this facilitation process. The system allows dynamic changes to datasets and organizational objectives for maintaining both usability and relevance of the system [13, 23].

ADCS demonstrated its workflow streamlining capacity in social enterprises by cutting down manual data processing by 35% more efficiently than traditional methods.

Comparison results between ADCS and traditional data-gathering methods show studies in Table 7 regarding time effectiveness and data precision along with scaling capabilities. Data processing with the ADCS system executes faster than traditional approaches because it cuts processing duration from 120 seconds to 30 seconds. The data accuracy achievement rate of ADCS stands at 90%, while traditional approaches only reach 70% due to their AI-enabled classification process. ADCS provides excellent flexibility, which allows it to accept large datasets, while traditional techniques struggle with scalability problems. The research confirms that ADCS develops data management excellence by implementing automated operations and intelligent processing methods.

Table 7
Comparative analysis of time efficiency, accuracy, and scalability: ADCS versus traditional methods

Criterion	ADCS	Traditional
Time Efficiency (seconds)	30	120
Data Accuracy (%)	90	70
Scalability	High	Low

The study simulated datasets along with case studies for validation because intra-cluster deployment constraints exist. The authors conducted sensitivity tests to measure performance fluctuations while modifying the database volume and chosen features together with cryptographic key durations. The ADCS system delivers reliability through consistent accuracy higher than 90% throughout different dataset arrangements. Added real-world tests would increase empirical evidence, but the research demonstrates proof of concept through its controlled evaluation systems. Researchers plan to conduct studies about the real-world deployment of the system with social enterprises for future work.

The findings of this study align with and extend previous research on automated data management systems and SDG alignment.

Automation and efficiency: Prior studies [3, 4] highlighted the benefits of automation in reducing manual data entry errors. ADCS builds on this by integrating iterative feedback mechanisms that dynamically refine keyword evaluations, a feature lacking in previous models.

Security integration: While Panagiotou et al. [5] demonstrated the effectiveness of cryptographic security in securing IoT applications, our study extends these principles to social enterprise data management, showing that AES-256 encryption and SHA-256 hashing can enhance data confidentiality and integrity.

Machine learning for SDG alignment: Baranowski [7] explored the use of LDA in thematic pattern recognition. Our results confirm that LDA, when combined with clustering algorithms, improves SDG classification accuracy by 32% over traditional frequency-based methods.

Multiple studies confirm the distinct value of ADCS since it unifies machine learning and security capabilities to resolve social enterprise challenges. The ADCS matches existing data processing security theories but needs additional evaluation to validate results. The AES-256 encryption procedure under ADCS along with RBAC strengthens data security but adds a processing time burden. The discovery opposes pre-existing beliefs about security improvements, which must not compromise system performance [6]. The future research should study optimization approaches to achieve a better security usability balance. LDA along with clustering models produced better alignment accuracy in global development goals, yet they also produced classification uncertainties when dealing with multidimensional datasets. Supervised learning techniques need to integrate with other methods for attaining better performance results. The effectiveness of ADCS for policy-based organizations might not automatically apply to commercial organizations with other data organization systems. Additional research needs to be conducted on a broader scale to determine general applicability.

Our study aligns with previous research demonstrating the effectiveness of AI and blockchain in securing large-scale systems.

Similar to the study by Alshammari [15], our ADCS utilizes machine learning for intelligent classification and blockchain for data protection. However, our research extends beyond intrusion detection by integrating RBAC and cryptographic hashing, which were not explicitly evaluated in previous frameworks. Furthermore, while Shinde et al. [16] focus on securing federated learning models using blockchain, our approach applies blockchain technology to SDG-aligned social enterprise data management, demonstrating a novel application of blockchain for mission-driven analytics. Additionally, studies such as [17, 18] highlight the impact of AI and blockchain in financial security, but their frameworks primarily target transaction verification and fraud detection, whereas our study emphasizes data authenticity and access control in social enterprises. These distinctions position our work as a unique contribution to data-driven, blockchain-secured data management for SDG alignment.

Several constraints exist regarding the application of this research, which merges AI and blockchain technology with secure data management practices for social enterprises. Security protocols that incorporate AES-256 along with SHA-256 hashing create computational overhead, which affects real-time system processing speed. The research should continue by analyzing either low-footprint encryption methods or optimization approaches to keep security compatibility with efficient system performance [18].

The research shows ADCS works well to protect social enterprise data while supporting SDGs, but the dataset evaluation includes only one group of social enterprises. The research findings show applicable results to various sectors, but certain implementation doubts exist. Validation of the framework needs various datasets to test it through organizational contexts covering commercial and governmental sectors [30].

The present execution of ADCS uses unsupervised learning frameworks for SDG alignment, but this method sometimes produces ambiguous classifications in complex datasets. Future research should study combination models that unite supervised learning through thematic mapping and unsupervised learning by clustering since these would boost classification precision and thematic mapping performance [31].

Similarly, while decentralization improves efficiency and control over the data in the system, it poses some issues related to the scalability of designs that revolve around the blockchain, particularly in terms of the time required to conduct a transaction and costs of storage space. Further research should explore layer 2 scaling solutions or adjust the consensus used in the creation of the proof-of-stake blockchain to minimize the discovered limitations while preserving the network's security and integrity [17]. Social enterprises need to thoroughly address and resolve regulatory as well as ethical obstacles when implementing AI and blockchain technologies. The research requires a shift toward studying legal frameworks because technology integration has become widespread across different sectors [32]. Research development targeting these restrictions will boost the effectiveness and expansiveness while promoting widespread adoption of data management systems that use blockchain technology and AI. They will successfully integrate into multiple demanding applications with secure intelligent data, which leads to meaningful, sustainable data governance after this advancement. ADCS delivers a new automated data-gathering framework that handles the core problems found in current literature reporting methods. The findings of this research contribute to both theoretical models and practical applications, which demonstrate to the data management literature about integrating automation and cryptographic security within one operational system. Previous research on machine learning for SDG alignment is extended through the addition of iterative feedback loops, which previous studies have not examined. The system presents an expandable, secure solution for social enterprises, which minimizes manual effort by 35% and maintains data authenticity and privacy preservation. This article presents a flexible ADCS model that organizations of all types, such as nonprofits and governmental agencies, along with the private sector working on sustainability initiatives, can customize to their needs. These contributions bring to light ADCS as an advanced system for automated data management, which brings both security and intelligence to organizations wanting to connect with SDGs. Future research must examine the deployment in actual environments and perform lasting measurements of results.

As the preceding discussion has demonstrated, the offered ADCS provides a powerful solution for organizations that are looking to achieve global sustainability while maintaining the flexibility and security of their data assets. The innovative and superior nature of the systems that are created from the sophisticated analytical and cryptographic approaches that are utilized serves as a standard for how future systems should be designed. In the future, the system may be enhanced to acquire more elements of predictive analysis for the purpose of anticipating an increase in the data relevant to the SDGs and expanding the spheres of application in which it can be used.

The ADCS demonstrates a significant contribution to the field of data analysis, as well as to the field of sustainability studies and policies. It has filled critical theoretical and empirical gaps in the existing body of literature as it could provide better network alignment and flexibility. It is demonstrated through the system's precise design that machine learning and cryptographic security can be used in conjunction with one another to address large-scale challenges that are consistent with the professional workspace.

5. Conclusion

The research developed in this study introduces a more pronounced change in data management of social enterprises, which features automation and data security in terms of enhancing operational efficiency and data integrity. It was found that ADCS significantly improves classification accuracy and processing speed, while making it a scalable solution also for organizations that want to synchronize their operations with SDGs. The ADCS has decreased manual processing time as compared with other data collection methods. The emphasis of the project is on enhancing both the efficiency and ensuring the safety of all information using cryptographic methods. ADCS also further increases the accuracy of SDG alignment by 32% over existing frequency-based classification methods. This is an adequate framework for enabling decision-making in social enterprises. It also minimizes human involvement in certain processes and their ability to make errors. Modern cryptographic techniques guarantee strong security assurances for privacy and authenticity. Furthermore, the machine learning application enables ADCS to categorize its initiatives according to certain SDGs. This improves the accuracy of thematic categories and supports more profitable strategic decision-making. This approach alleviates the manual workload for social enterprises and circumvents data management issues. Nonetheless, ADCS demonstrates a robust data management system. However, it is essential to enhance cryptographic security protocols as they may impose computational demands that could expose domain responsiveness. Furthermore, employing supervised machine learning techniques can enhance data categorization as the model significantly boosts classification accuracy. Future research should focus on how ADCS can be used in a wide range of AI-driven situations to make sure it works in all kinds of operational situations along with improvements like predictive analytics and real-world adaptability.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request from the corresponding author.

Author Contribution Statement

Fardin Muttaki: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization, Project administration. **Aqeel Sahi:** Conceptualization, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision. **Shahab Abdulla:** Methodology, Validation, Investigation, Data curation, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Kaled Aljebur:** Software, Resources.

References

[1] Créquit, P., Boutron, I., Meerpohl, J., Williams, H. C., Craig, J., & Ravaud, P. (2020). Future of evidence ecosystem series: 2. current opportunities and need for better tools and methods. *Journal of Clinical Epidemiology*, *123*, 143–152. https://doi.org/10.1016/j.jclinepi.2020.01.023

[2] Tlouyamma, J., & Mokwena, S. (2024). Automated data quality control system in health and demographic surveillance system. *Science, Engineering and Technology*, *4*(2), 82–91. https://doi.org/10.54327/set2024/v4.i2.125

[3] Nilashi, M., Ooi, K. B., Tan, G., Lin, B., & Abumalloh, R. (2023). Critical data challenges in measuring the performance of sustainable development goals: Solutions and the role of big-data analytics. *Harvard Data Science Review*, *5*(3). https://doi.org/10.1162/99608f92.545db2cf

[4] Zhao, W., Aldyaflah, I. M., Gangwani, P., Joshi, S., Upadhyay, H., & Lagos, L. (2023). A blockchain-facilitated secure sensing data processing and logging system. *IEEE Access*, *11*, 21712–21728. https://doi.org/10.1109/ACCESS.2023.3252030

[5] Panagiotou, P., Sklavos, N., Darra, E., & Zaharakis, I. D. (2020). Cryptographic system for data applications, in the context of internet of things. *Microprocessors and Microsystems*, *72*, 102921. https://doi.org/10.1016/j.micpro.2019.102921

[6] Ferencz, K., & Rigó, E. (2022). Analysis of network-based data collection mechanisms applicable in securing critical infrastructures. In *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics*, 299–304. https://doi.org/10.1109/CINTI-MACRo57952.2022.10029629

[7] Baranowski, M. (2022). Epistemological aspect of topic modelling in the social sciences: Latent Dirichlet allocation. *Przegląd Krytyczny*, *4*(1), 7–16. https://doi.org/10.14746/pk.2022.4.1.1

[8] Aly, E., Elsawah, S., & Ryan, M. J. (2022). A review and catalogue to the use of models in enabling the achievement of sustainable development goals (SDG). *Journal of Cleaner Production*, *340*, 130803. https://doi.org/10.1016/j.jclepro.2022.130803

[9] Wu, B., Tian, F., Zhang, M., Zeng, H., & Zeng, Y. (2020). Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. *Geography and Sustainability*, *1*(1), 25–32. https://doi.org/10.1016/j.geosus.2020.03.006

[10] Almaqbali, I. S. H., Al Khufairi, F. M. A., Khan, M. S., Bhat, A. Z., & Ahmed, I. (2020). Web scrapping: Data extraction from websites. *Journal of Student Research*, 1–4. https://doi.org/10.47611/jsr.vi.942

[11] Nigam, H., & Biswas, P. (2021). Web scraping: From tools to related legislation and implementation using Python. In *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020,* 149–164. https://doi.org/10.1007/978-981-15-9651-3_13

[12] Kbean, N. A. W., & Sadkhan, S. B. (2020). Cryptography techniques within SCADA system-A survey. In *2020 3rd International Conference on Engineering Technology and its Applications*, 89–94. https://doi.org/10.1109/IICETA50496.2020.9318832

[13] Hoyas, S. (2022). ASDG—An AI-based framework for automatic classification of impact on the SDGs. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, 119–123. https://doi.org/10.1145/3560107.3560128

[14] Fadi, O., Karim, Z., Abdellatif, E. G., & Mohammed, B. (2022). A survey on blockchain and artificial intelligence technologies for enhancing security and privacy in smart environments. *IEEE Access*, *10*, 93168–93186. https://doi.org/10.1109/ACCESS.2022.3203568

[15] Alshammari, B. M. (2023). AIBPSF-IoMT: Artificial intelligence and blockchain-based predictive security framework for IoMT technologies. *Electronics*, *12*(23), 4806. https://doi.org/10.3390/electronics12234806

[16] Shinde, R., Patil, S., Kotecha, K., & Ruikar, K. (2021). Blockchain for securing AI applications and open innovations. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(3), 189. https://doi.org/10.3390/joitmc7030189

[17] Rane, N., Choudhary, S., & Rane, J. (2023). Blockchain and artificial intelligence (AI) integration for revolutionizing security and transparency in finance. *SSRN*. https://doi.org/10.2139/ssrn.4644253

[18] Chowdhury, R. H. (2024). Blockchain and AI: Driving the future of data security and business intelligence. *World Journal of Advanced Research and Reviews*, *23*(1), 2559–2570. https://doi.org/10.30574/wjarr.2024.23.1.2273

[19] Ganz, F., Puschmann, D., Barnaghi, P., & Carrez, F. (2015). A practical evaluation of information processing and abstraction techniques for the internet of things. *IEEE Internet of Things Journal*, *2*(4), 340–354. https://doi.org/10.1109/JIOT.2015.2411227

[20] Gopal, A. V., & Ramanathan, K. (2022). Automating data collection process in industry 4.0: A user case study. In *AIP Conference Proceedings*, *2519*, 030027. https://doi.org/10.1063/5.0109812

[21] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, *13*(3), 144–168.

[22] Bhatt, C., Gaitri., Kumar, D., Chauhan, R., Vishvakarma, A., & Singh, T. (2023). Web scraping: Huge data collection from web. In *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology*, 375–378. https://doi.org/10.1109/ICSEIET58677.2023.10303037

[23] Oladejo, B., & Hadžidedić, S. (2021). Electronic records management–A state of the art review. *Records Management Journal*, *31*(1), 74–88. https://doi.org/10.1108/RMJ-09-2019-0059

[24] Rashid, J., Shah, S. M. A., & Irtaza, A. (2019). Fuzzy topic modeling approach for text mining over short text. *Information Processing & Management, 56*(6), 102060. https://doi.org/10.1016/j.ipm.2019.102060

[25] Sarkar, A., Chatterjee, S. R., & Chakraborty, M. (2021). Role of cryptography in network security. In M. Chakraborty, M. Singh, V. E. Balas, & I. Mukhopadhyay (Eds.), *The "essence" of network security: An end-to-end panorama* (pp. 103–143). Springer. https://doi.org/10.1007/978-981-15-9317-8_5

[26] Kavitha, A., Rao, B. S., Akthar, N., Rafi, S. M., Singh, P., Das, S., & Manikandan, G. (2022). A novel algorithm to secure data in new generation health care system from cyber attacks using IoT. *International Journal of Electrical and Electronics Research*, *10*(2), 270–275. https://doi.org/10.37391/IJEER.100236

[27] Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. *Sustainability*, *14*(5), 2497. https://doi.org/10.3390/su14052497

[28] Calle-Arango, L. F., la Cruz, A., Santiago, M., & Pérez, A. (2022). Big data techniques applied to stakeholder data for non-governmental organizations. In *2022 IEEE Sixth Ecuador Technical Chapters Meeting*, 1–6. https://doi.org/10.1109/ETCM56276.2022.9935756

[29] Ranabahu, N. (2020). 'Wicked' solutions for 'wicked' problems: Responsible innovations in social enterprises for sustainable development. *Journal of Management & Organization*, *26*(6), 995–1013. https://doi.org/10.1017/jmo.2020.20

[30] Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., & Soyombo, O. T. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. *International Journal of Science and Research Archive*, *11*(1), 2101–2110. https://doi.org/10.30574/ijsra.2024.11.1.0279

[31] Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology*, *67*(10), 2464–2476. https://doi.org/10.1002/asi.23596

[32] Elkhodr, M., Khan, S., & Gide, E. (2024). A novel semantic IoT middleware for secure data management: Blockchain and AI-driven context awareness. *Future Internet*, *16*(1), 22. https://doi.org/10.3390/fi16010022

How to Cite: Muttaki, F., Sahi, A., Abdulla, S., & Aljebur, K. (2025). An Enhanced Data Collection System for Social Enterprises: Securing Impact with Machine Learning and Cryptography. *FinTech and Sustainable Innovation*, *1*, A11. https://doi.org/10.47852/bonviewFSI52025209