CLOUD DATA SECURITY AUDIT REPORT TECHNIQUES USING BAT INSPIRED ALGORITHM: A REVIEW

ONUJA, A. M.[1], ADEBAYO, O. S[2], OLALERE, M.[3], ALIYU, H. O.[4]

[1,2,4]Federal University of Technology, Minna, Nigeria.

[3]Islamic University, Uganda

Correspondence: onuja14@gmail.com

Abstract

Cloud computing has become an interesting attraction for Information and Communication Technology (ICT) practitioners. It is helping to reduce the cost of procuring computing devices and services. To this end, security attacks have been rising with cloud service providers always having to under-report the prevalence in order not to scare away potential cloud service subscribers. In this review, eighty (80) papers were acquired from reputable academic publications using Google Scholar as the search engine. The eighteen (18) carefully selected from the pool of 80, for review, have a total of seventy-four (74) versions on the internet. Furthermore, the authors agree with the research publications, that there is need to bring on board the client into security breach mitigation strategies, being developed to strengthen cloud data security. Findings from this review, reveals that research interest in cloud data security audit has increased from four (4) papers between 2011 and 2016, to thirteen (13) papers between 2017 and 2022. These existing frameworks and techniques are still being tested to improve performance. This research, therefore, proposed a cloud data security audit report technique using Bat inspired algorithm (CDSART-BA) to improve cloud data security.

Keywords: *Audit Report, Bat Algorithm, Cloud computing, Data, Security, Smart Environment.*

Introduction and Background

Smart environments are gradually gaining research attention and trends with respect to adoption and usability in many human endeavours (Adebesin et al., 2021). Current developments in technology have paved the way for increased safety, security and at the same time increased user experience (Ikuomola and Bashir, 2021) and privacy threats. The major threat to the electronic banking system and other services that support smart nations are the concerns of security and privacy of information (Umoren et al 2021). Cloud computing services is a modern trend in information and communication technology (Ngumbi and Wasike, 2022) that is evolving. It is powered with high-speed internet connection. Commonly used resources like hardware, system and user applications can be sourced as service, over the internet (Shankarwar and Pawar, 2015). Figure 1 one below, presents a diagram of cloud computing.

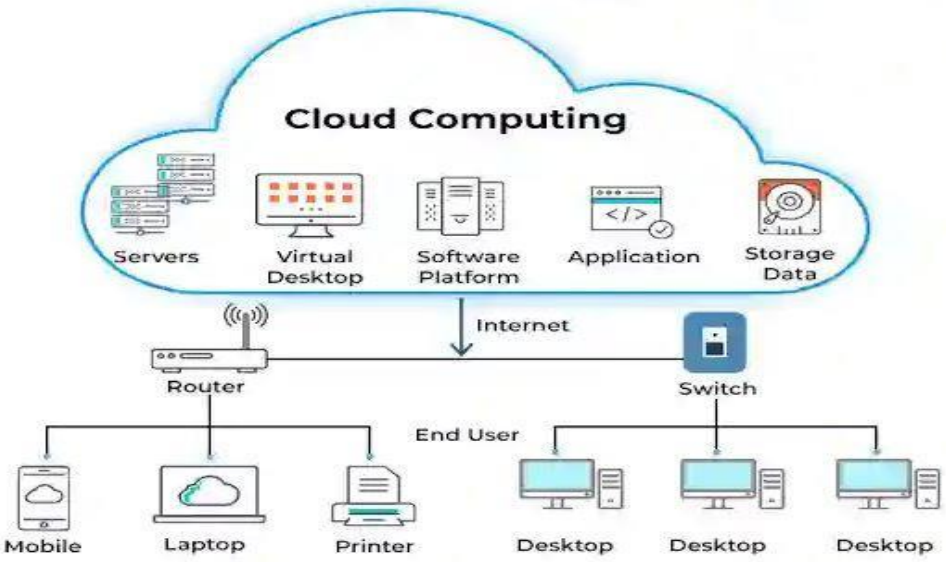

Figure 1: Cloud computing (Fahad, Ahmed and Kahar, 2018). [6].

Cloud computing services have markets that are growing fast, but mainly, security reasons restrict broad wide industrial acceptance (Doelitzscher, 2012). The high number of research with financial support, is a reflection of the expanding cloud security concerns. An example is the European Union 7th Framework Projects, whose objective 1.2 is centered on internet of services (Doelitzscher, 2012). Germany included a trusted cloud, as part of the government 's economic development plan (Doelitzscher, 2012). In this concept, Cloud users must adopt their security, backup and business continuity plans to avoid losing valuable data in the cloud (Doelitzscher, 2012).

Enron and WorldCom financial scandals raised concerns by the government about accounting errors and fraudulent practices created within organizations (Fahad et al. 2018). The Sarbanes–Oxley Act (SOX) of 2002 was legislated (Fahad et al. 2018) and Governments policy for Information Security is recommending security audits and certificates as the preferred method of proof to clients of cloud services (Doelitzscher, 2012). IT auditing thus becomes mainstream practice (Fahad, et al. 2018). Figure 2 below is a diagram identifying the role of auditing in cloud data security and accountability life cycle.

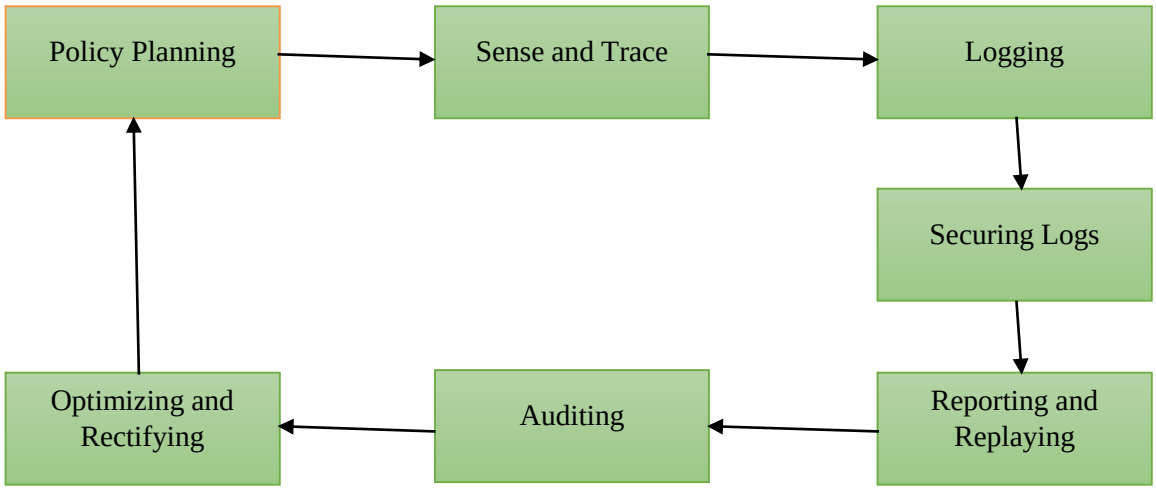

Figure 2: Auditing in Cloud data security accountability life cycle (Ko et al., 2018).

Related Work

Many researchers have actually worked on cloud data security auditing with different techniques and trade off. This serves a basis for this research work aimed to improve on the findings of the papers surveyed.

A. Cloud data security

In supporting the aim of this research, Ko et al., (2018) agrees that preventive controls for security and privacy measures are actively being researched and that there is still little focus on detective controls related to cloud accountability and auditability. The paper presents the TrustCloud framework as a solution but the overhead cost was not highlighted. Many researches have been conducted with the help of Third-Party Auditors (TPA), to check stored data in the cloud to assure the integrity of the environment (El-Booz, Attiya, and El-Fishawy, 2016). Since the cloud service provider (CSP) has the monopoly of interaction with TPA, the CSP most times, is able to influence the concealment of data loss, from cloud subscribers, to prevent defamation (El-Booz et al 2016). The proposed system enhances the authentication level of security by using two authentication techniques; time-based one-time password (TOTP), for cloud users' verification and automatic blocker protocol (ABP), to fully protect the system from unauthorized third-party auditors. However, the mechanism by which auditing reports are generated was not discussed. The proposal in (Ismail and Islam, 2020), is a framework for cloud security transparency, and audit in addition to developing a tool, through which users can collect and analyze evidences from cloud service providers, for determining conformity to requirements, as well as for the specification of remedial actions. The concept of cloud data auditing system in (Mohan and Gladston, 2020) integrates MerkleTree based Cloud audit together with blockchain based audit recording system, thereby recording each verification result into a blockchain as a transaction with a timestamp. This paper suggests the use of the load balancing ability of Bat algorithm to generate audit reports with lower overhead cost.

B. Application of bat algorithm

Bat algorithm (Valarmathi and Sheela, 2019) is used to improve optimization for scheduling of tasks in a cloud environment. The proposed model in (Sreeram and Vuppala, 2019) is the bat algorithm. It is applied as a technique to achieve prompt and fast detection (Mutheu and Wasike, 2020) of distributed denial of service, in the application layer, through HTTP flooding. This justifies the view of this research that bio-inspired bat algorithms can be used as a technique for cloud data security audit reports. Similarly, (Raja Sree, and Mary Saira Bhanu, 2020) present fuzzy bat clustering techniques that read activity logs in cloud virtual machines, and investigates HTTP flooding attacks by grouping identical input patterns. It then determines the anomalous behaviour using deviated anomalous scores.

Sagnika et al (2018) presents an application of the Bat Algorithm (BA) for workflow schedules. The algorithm is successful in laboratory implementation, gives an optimal processing cost, convergence, fair load balancing versus Cat Swarm Optimization (CSO) and Particle Swarm Optimization (PSO).

The discussion in (Ullah et al 2020) is concerned with virtual machines as a tool for resource distribution. This is because when users send data into virtual machines, it can be overloaded sometimes, thus leading to breakdown, thereby affecting data security goals. The researchers used the bat algorithm as a load balancing technique to improve and enhance even the resource allocation system for cloud virtual machines (VM). Bat algorithm also has echolocation potentials that enables it to send auditing reports to users as proposed by this research.

Punitha and Indumathi, (2019) submitted that data integrity protection becomes difficult, as users' subscriptions to cloud computing services increase. There is therefore a need for clients to become part of the security control process with increasing security threats. They suggested accountability with a key generation framework, to maintain users' data in the cloud. This would encircle the logging mechanism together with users' data and policies.

However, bat key generation will require more computing overhead cost hence this paper proposes that the bat algorithm should rather just send an audit report to client to take necessary security measures.

Proposed Method

A. Survey of Research Problems

The research applies the method used in (Yakubu et al., 2019) to carry out the survey for the study. Eighty (80) papers were downloaded between January and July, 2022. The source of the 17 papers cited or reviewed for this work are Nigeria Computer Society library archive, Springer, IEEE, EURASIP, ACM, Elsevier, Research Gate. Notably, Google Scholar served as the gateway to all these research databases. The remaining 18 were sorted from a total of 74 versions. The time range or year interval for the publication used in the review is between 2011 and 2021 as grouped with a class interval of 3 years in table 1 below.

Table 1: Publication Year Interval

2011-2013	[5], [7].	2
2014-2016	[4], [8].	2
2017-2019	[6], [11], [12], [14], [16], [17].	6
2020-2022	[1], [3], [9], [10], [13], [15], [18], [19].	9
Total		18

B. Security Audit Report Model

This research paper is proposing a methodology of using bat algorithms as a technique for cloud data security reports for clients using cloud services. Hitherto, cloud data security audit has been the sole privilege of the cloud service provider (CSP). This was to avoid dissent in patronage. However, current research in cloud computing is bearing the fact that cloud computing clients have to be active partners in the security of data in the cloud. So, thus, there is security requirement for cloud data security audit reports to include clients. Existing research models are shown in figure 3, in which cloud service organization's administrators grant access to cloud users to use their platform as the cloud service provider. There is synchronization between the CSP, administrative and operational third-party auditor, while the client is isolated from the auditor. The model in figure 3 looks at security auditing from the perspective of authentication with CSP deciding what eventual security breach report it relays to the client.

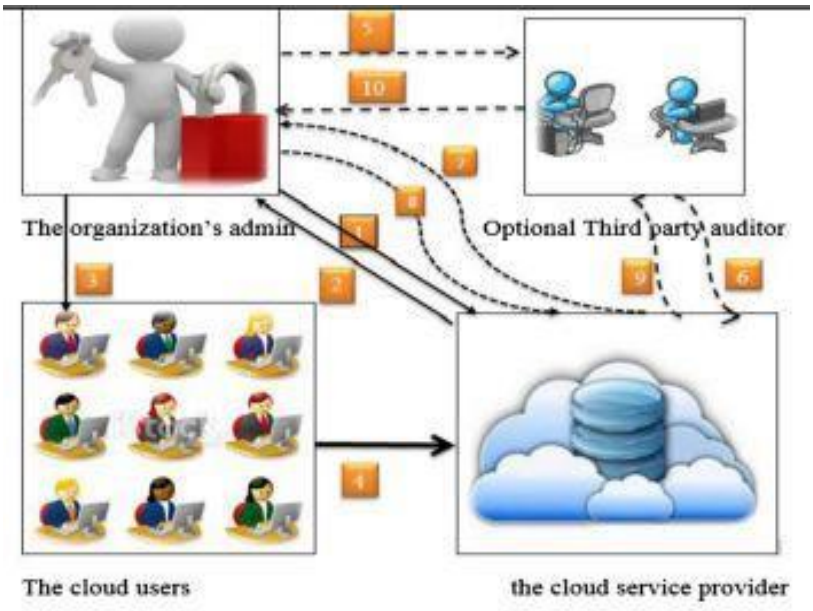

Figure 3: Existing security audit model (El-Booz, Attiya, and El-Fishawy, 2016) [8].

C. CDSART-BA Workflow

Given that cloud service clients have signed into the model cloud environment proposed in this paper, the embedded bat algorithm keeps track of all data transiting from the client to the cloud server. The BA algorithm protects against data loss and flooding of traffic into the communication channel. If the size of source packets of data in bits from the client is equal to the size of packets arriving in the cloud server, the CDSART-BA model sends a secured audit report to the client through email. Also, if an addition or subtraction is noticed in the packets of data in bits from client side to cloud server end, an audit report signaling security threat is sent, enabling the client to take necessary security actions in collaboration with the cloud service provider (CSP). Figure 4 presents an improved technique on the existing model, that gives direct security report access from the cloud service provider to the cloud service client (CSC)

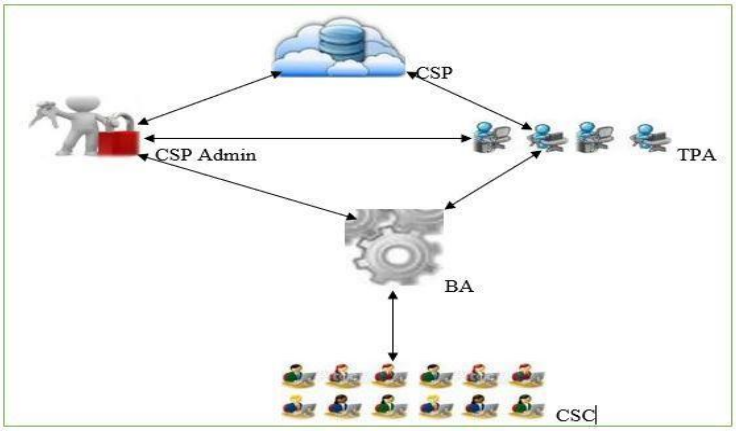

Figure 4: CDSART-BA Model

D. CDSART-BA Mathematical Model

Mathematically, let the initial client data cd_0, the initial stored cloud data sd_0, and the subsequent client data cd_n and subsequently stored cloud data sd_n be variables for auditing security and transaction between client or cloud service user and cloud service provider. And given that, the number of packets transmitted is greater or equal to zero.

Security is achieved when;

$$\sum_{i=0}^{n} cd = \sum_{i=0}^{n} sd \ldots\ldots\ldots\ldots\ldots\ldots equation\ 1$$

Security is breached when;

$$\sum_{i=0}^{n} cd \leq \sum_{i=0}^{n} sd \ldots\ldots\ldots\ldots\ldots\ldots equation\ 2$$

Security is also breached when;

$$\sum_{i=0}^{n} cd \geq \sum_{i=0}^{n} sd \ldots\ldots\ldots\ldots\ldots\ldots equation\ 3$$

The equation above shows three scenarios that trigger a security audit report to the client. The equation one (1) describes a good security report. Equation two (2) describes data loss scenario and equation three (3) explains data intrusion and unauthorized addition while on transit.

Discussion of Survey Outcomes and Results

This review employs use of google scholar as a gateway into academic research databases, for the survey of resourceful research publications.

Existing techniques include; time-based one-time password (TB-OTP) and automatic blocker protocol (ABP) (El-Booz et al 2016). The method used in (El-Booz et al 2016) secures client data against the access of CSPs and TPAs without authorization using TB-OTP and ABP. So, an audit report cannot be generated without the permission of the client. Sending permission requests to individual cloud users improves security of client data but increases overhead.

Merkle Tree and Blockchain based cloud data (Mohan and Gladston, 2020) effects timespan and enables cloud users to verify that TPAs performed their auditing services as at when due. Centralized cloud information accountability with bat key generation algorithm (CCIA-BKGA) framework in cloud computing environment (Punitha and Indumathi, (2019). This

approach applies use of the bat algorithm to generate encryption keys to authenticate and account for access into cloud user data thereby adding to processing requirements of computing devices.

In contrast, the proposed CDART-BA uses a bat algorithm to trigger audit reports to cloud users, CSPs and TPAs, for an all-inclusive security mitigation strategy. The summary of survey outcomes and results is presented on table 2.

Table 2: Survey outcomes and Results

1.	Delitzscher et al., (2012)	IEEE	Validating cloud infrastructure changes by cloud audits.	The model is for Service provider-side audit reports.
2.	Ko et al., (2011)	IEEE	TrustCloud: A framework for accountability and trust in cloud computing.	Service provider oriented.
3.	Shankarwar and Pawar, (2015).	Springer	Security and privacy in cloud computing: A survey.	Research findings are yet to be implemented.
4.	El-Booz, Attiya, and El-Fishawy (2016).	EURASIP	Automatic blocker protocol and one-time password for	Did not affect the audit report.

			secure cloud storage.	
5.	**Fahad, Ahmed** and Kahar (2018).	Springer	Combines artificial neural network and whale swarm to detect intrusion cloud computing.	The focus of the research is detection and not audit reports for improved security strategies.
6.	Valarmathi, and Sheela, (2019).	Springer	Applies bat algorithm and particle swarm optimization for task scheduling in cloud environments.	Basically about the application of the bat algorithm and did not utilize the algorithms for the audit report.
7.	Sreeram and Vuppala**,** (2019).	Elsevier	HTTP flood attack detection in application layer using machine learning metrics and bat inspired algorithm for the	The research did not solve the audit report requirement.

			detection of HTTP flood attacks.	
8.	**Sagnika, Bilgaiyan** and Mishra, (2018).	Springer	Scheduling of data workflow in cloud computing using bat algorithm.	The aim of the paper is about load balancing and not audit report.
9.	**Punitha** and Indumathi, (2019).	Springer	The paper centralized the authentication of information with keys generated using the bat algorithm.	The implementation of the framework.
10.	**Yakubu et al.,** (2019).	Springer	Security challenges in fog-computing environments.	Developing a model that improves the security in fog-computing.
11.	**Adebesin et al.,** (2021).	Nigeria Computer Society	A framework for mitigating phishing attacks.	The implementation of the framework in a cloud environment.

12.	Umoren, Inyang, **and Gilean** (2021).	Nigeria Computer Society	Security threats prediction using Bayesian network algorithm in banks digital channels.	Developing the system for cloud environments.
13.	Ismail and Islam, (2020).	Nigeria Computer Society.	Framework for cloud security and audit.	Authors do not specify the security algorithm proposed in the framework.
14.	Mohan **and Gladston,** (2020).	ACM	Merkle tree and Blockchain-based cloud data auditing.	The use of blockchain technology has associated high computational power requirements.
15.	Raja Sree, and Mary Saira Bhanu, (2020).	Springer	Detection of HTTP flooding attacks in the cloud using fuzzy bat clustering.	Affecting audit report in the model for improved security.

16.	Ullah, **Nawi** and Khan, (2020).	INDERSCIENCE	Load balancing using bat algorithm.	Development of the system model.
17.	**Ngumbi** and Wasike, (2022).	Kirinyaga University	ICT use in teaching and learning.	Improving the security of cloud environments for teaching and learning.
18.	Mutheu and Wasike, (2020).	Kirinyaga University	Decision Support (IDS) in Software Risk Management Based on Data Mining, Rough Sets and Decision Theory.	Applying decision theory in cloud data audit reports.

Conclusion

In this review, the authors were able to contribute to research on cloud data security with respect to auditing reports that involve cloud service clients as important stakeholders in the quest for the confidentiality, integrity and availability of cloud data and services. The review is able to achieve system design and mathematical models from the literature review of related work. The authors look forward to implementing the prototype of cloud data security report techniques using bat algorithm (CDSART-BA), to send timely audit reports to cloud users about the security situation of their data online. This is hoped to be achieved in Google Firebase, a google cloud computing lab environment that supports the use of backend codes.

References

1. Adebesin A.A, Sodiya A.S, Orunsolu A.A, Lawal O.A, Kareem S.O. (2021). A Proposed Distributed Anti-Phishing Framework for mitigating Cyber-attacks in Smart Environments. International Journal of Information Security, Privacy on Digital Forensics. 5(2). Nigeria Computer Society
2. Doelitzscher F., Fischer C., Moskal D., Reich C., Knahl M., Clarke N. (2012). Validating cloud infrastructure changes by cloud audits. In *2012* IEEE Eighth World Congress on Services. 377-384. IEEE.
3. El-Booz S. A., Attiya G., El-Fishawy N. (2016). A secure cloud storage system combining time-based one-time password and automatic blocker protocol. EURASIP Journal on Information Security. (1) 1-13.
4. Fahad A. M., Ahmed A. A., Kahar M. N. M. (2018). Network intrusion detection framework based on whale swarm algorithm and artificial neural network in cloud computing. In International Conference on Intelligent Computing and Optimization 56-65. Springer, Cham.
5. Ismail U. M., Islam S. (2020). A unified framework for cloud security transparency and audit. Journal of Information Security and Applications, 54, 102594. Nigeria Computer Society.
6. Mohan A. P., Gladstone A. (2020). Merkle tree and Blockchain-based cloud data auditing. International Journal of Cloud Applications and Computing (IJCAC). 10(3): 54-66. ACM.
7. Mutheu R., Wasike J. (2020). Intelligent Decision Support (IDS) in Software Risk Management Based on Data Mining, Rough Sets and Decision Theory. African Journal of Science, Technology and Engineering (AJSTE), *1*(1):82-96. Kirinyaga University.
8. Ngumbi E., Wasike J. (2022). Use of ICT in the Classroom: The Teacher and Learner. African Journal of Science, Technology and Engineering (AJSTE). *3*(1):34-48. Kirinyaga University.
9. Punitha A., Indumathi G. (2019). Centralized cloud information accountability with bat key generation algorithm (CCIA-BKGA) framework in cloud computing environment. Cluster Computing, 22(2):3153-3164. Springer.

10. Sagnika, S., Bilgaiyan, S., Mishra, B. S. P. (2018). Workflow scheduling in a cloud computing environment using the bat algorithm. In Proceedings of first international conference on smart system, innovations and computing Springer, Singapore. (149-163).

11. Shankarwar M. U., Pawar A. V. (2015). Security and privacy in cloud computing: A survey. In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) (1-11). Springer, Cham.
12. Sree T.R., Bhan, S.M. (2020). Detection of HTTP flooding attacks in the cloud using fuzzy bat clustering. Neural Computing and Applications, 32(13):9603-9619. Springer.

13. Sreeram I., Vuppala V. P. K. (2019). HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm. Applied computing and informatics, 15(1): 59-66. Elsevier

14. Ullah, A., Nawi, N. M., and Khan, M. H. (2020). BAT algorithm used for load balancing purpose in cloud computing: an overview. International Journal of High Performance Computing and Networking, 16(1):43-54. INDERSCIENCE

15. Umoren I., Inyang S., Gilean O. (2021). Bayesian Network Algorithm for Predictive Modelling of Cyber Security for Efficient Bank Channels Digitalization. International Journal of Information Security, Privacy on Digital Forensics 5(2).

16. Valarmathi R., Sheela T. (2019). Ranging and tuning based particle swarm optimization with a bat algorithm for task scheduling in cloud computing. Cluster Computing, *22*(5): 11975-11988.

17. Yakubu, J., Abdulhamid, S. I. M., Christopher, H. A., Chiroma, H., and Abdullahi, M. (2019). Security challenges in fog-computing environment: a systematic appraisal of current developments. Journal of Reliable Intelligent Environments, 5(4): 20

BIG DATA ANALYTICS FOR SMES' PERFORMANCE SUSTAINABILITY IN POST-COVID-19 KENYA

Musyoka C., Wanjohi. P.

Department of Pure and Applied Sciences

Kirinyaga University, Kenya

Correspondences: cmusyoka@kyu.ac.ke

Abstract

Small and medium-sized businesses (SMEs) play a critical role in a nation's economy, contributing significantly to its wealth and fostering innovation. Globally, they account for up to half of all jobs and 90% of all businesses. However, SMEs often grapple with limited access to credit from suppliers, compounded by liquidity challenges, decreased sales, and defaults, as supply chains struggle to secure credit. Factors like advance payments, penalties for delayed credit payments, and recurring expenses further exacerbate the vulnerability of Smashing the context of Kenya's post-COVID-19 landscape. This study aimed to explore the potential of Big Data Analytics and Data Science in sustaining SMEs' performance. Drawing from theories such as Complex Adaptive System and Strategic Choice Theory, a descriptive survey design was employed, on a target population of 287 managers of SMEs in each sub County, in Nairobi. Employing a stratified sampling method, a total of 260 respondents were interviewed. Data was analyzed using descriptive statistics, including frequencies, percentages, mean, and standard deviation, while inferential statistics like multiple regression and Pearson correlation were used to examine relationships between variables. The study revealed that Business Intelligence, with a mean score of 3.9 (std. dv = 0.851), and Machine Learning, with a mean of 3.7 (std. dv = 0.928), both had a positive impact on SMEs' sustainability, with an overall average mean of 3.8 (std. dv = 0.8895). Similarly, Data Analytics, comprising Predictive Analytics (mean = 3.73, std. dv = 0.850) and Prescriptive Analytics (mean = 3.85, std. dv = 0.684), positively influenced SMEs' performance, with an average mean of 3.79 (std. dv = 0.767). These findings underscore the potential of Data Science drivers like Business Intelligence and Machine Learning in helping SMEs tackle unforeseen challenges in competitiveness. The study further highlights the importance of implementing a robust legal framework to safeguard data in the context of Data Analytics, particularly in predictive and prescriptive analysis, as a means to enhancing SMEs' performance, survival, and growth in the post-COVID-19 era.

Keywords: *Big Data Analytics, Data Analytics, Business Intelligence, Machine Learning, Predictive Analytics*

1.0 Introduction

Small and medium-sized businesses (SMEs) contribute to approximately half of all jobs and 90% of businesses worldwide but were significantly affected globally due to COVID 19 pandemic because of their small size and limited resources (Kumar *et al.*, 2021). They continue to face limited credit availability from creditors, who also face liquidity problems, lower sales, and higher rates subjecting the sector to severe financial constants. The sector continues to face severe financial constraints as a result of supply chains not receiving credit from suppliers. A decrease in household income of approximately 36% was experienced by 95% of the entrepreneurs. Household income continues to be impacted by lower sales and family members losing jobs or seeing their wages reduced.

According to KBA MSMEs Survey Report, (2021) 35% of them missed or delayed loan repayments. Thus more small and medium-sized enterprises (SMEs) required financing to continue operating yet business owners struggle to obtain the necessary liquidity to revive and flourish. The study indicated that the majority of merchants accepted digital payments before the pandemic. However, a significant increase in the use of digital financial services resulted from exclusion of transactions exceeding 1,000 Ksh. (10 USD). Thus between February and October 2020, the Central Bank of Kenya reported that the monthly volume of P2P transactions increased by 87 percent. The volume of transactions below Ksh. 1,000 (USD10) increased by 114 percent during this time, and 2.8 million more customers began using mobile money, resulting in significant growth in digital business transactions. A slight improvement of 8% was observed among SMEs that use partnerships or an e-commerce model to mitigate the negative effects of COVID-19. More SMEs in rural areas—89 percent compared to 62 percent in urban areas have yet to adopt any digital channel to grow their businesses and adapt to the shifting business environment (KBA MSMEs Survey Report, 2021).

Majority of SMEs are microfinance institution customers who have drained their reserve funds and offered resources to relieve the extreme monetary effect of the pandemic but still ran the risk of having their credit limits reduced as their savings decrease, making them more vulnerable and less able to handle unexpected events in the future. In order to keep their businesses afloat, SMEs have been learning how to diversify their sources of income (KBA MSMEs Survey Report, 2021). Thus, have been employing a variety of

strategies to stay afloat. These strategies include seeking credit and using loan proceeds to infuse funds into their businesses while also prioritizing business continuity and seeking external support to recover. Vitari, and Raguseo (2020) stated that Data Analytics will help SMEs to survive in today's dynamic business environment based on Predictive Analytics and Prescriptive Data Analytics to enable them make more data-driven decisions based on historical and present data and to predict future business dynamics and respond as appropriate.

Statement of the problem

In the wake of the COVID-19 pandemic, Kenya's small and medium-sized enterprises (SMEs) faced formidable obstacles in their quest for sustained performance and long-term viability. The pandemic has brought havoc on traditional business models and consumer behavior, necessitating a rapid adaptation of SMEs to the ever-changing business landscape. However, a significant impediment lies in the fact that many SMEs in Kenya lack the essential tools and strategies required to harness the potential of big data analytics effectively. This deficiency hampers their ability to make data-driven decisions that could bolster their performance and ensure sustainability in a post-pandemic world (Chumba *et.al*., 2020).

Small and Medium Enterprises (SMEs) face a critical challenge in that their ability to navigate unforeseen situations is hindered by limited innovative technological capacities and underdeveloped research and development capabilities (Nath and Agrawal, 2020). These challenges become particularly pronounced during and post the COVID-19 pandemic. Survivability for SMEs therefore hinges on their capacity to address these problems by fortifying technological capabilities, especially in customer relations management (CRM). Additionally, there is a pressing need for SMEs to establish a flexible work system, allowing employees to work seamlessly from any location at any time. Failure to address these technological shortcomings poses a significant obstacle to SMEs' adaptability and sustainability in the face of ongoing and unforeseen challenges.

These challenges are compounded by various factors, including financial constraints, limited access to advanced technology, and lack of awareness and expertise in the realm of big data analytics. Additionally, SMEs in Kenya are operating in an intensely competitive environment, where adaptability and responsiveness to shifting market

dynamics are critical for survival. Without the capacity to leverage big data analytics, SMEs may struggle to identify emerging trends, understand customer preferences, and optimize their operational processes, potentially leaving them disadvantaged

Objective of the study

To establish the extent to which Big Data Analytics influence sustainability and performance improvement of small and medium-sized enterprises (SMEs) in Kenya after COVID-19 pandemic.

Research Question

The research was guided by the following research question:

To what extent does applications of Big Data Analytics influence the sustainability and performance improvement of small and medium-sized enterprises (SMEs) in Kenya the Post-COVID-19 pandemic?

Empirical Review

Maroufkhani et al. (2020) provides compelling insights into the transformative potential of BDA on SME sustainability and performance. The study effectively highlights how leveraging extensive and diverse datasets empowers SMEs to gain valuable insights into customer behaviors, identify market trends, and optimize operational efficiency. The emphasis on data-driven decision-making and strategic resource allocation underscores BDA's role as a catalyst for overall business improvement. The recognition of real-time analysis as a tool for proactive problem-solving and uncovering growth opportunities is commendable. Given the challenges posed by the post-COVID-19 era, the study rightly emphasizes BDA's contribution to equipping SMEs with the resilience and agility required to navigate dynamic market conditions. However, a critical examination of the study should consider the specifics of the proposed adoption model, potential limitations, and the generalizability of findings across different SME contexts to ensure the robustness and applicability of the proposed insights. Additionally, a deeper exploration of the challenges and ethical considerations associated with BDA adoption in SMEs would contribute to a more comprehensive understanding of its implications.

According to Nasrollahi *et al.,* (2021) study investigates the impact of big data adoption (BDA) on the performance of small and medium enterprises

(SMEs), presenting a comprehensive model. The research unveils BDA's multifaceted influence on SMEs, positively affecting both operational (efficiency, productivity) and economic (profitability, revenue) performance. Notably, no direct association was found between BDA and social performance. The study highlighted the mediating role of operational performance, where improvements in efficiency due to BDA contributed to enhanced economic performance. The model is conceptualized as a pyramid, with BDA components forming the base and arrows indicating causal relationships between layers. The findings emphasize BDA's potential for SMEs to streamline operations, make data-driven decisions, and gain a competitive edge. While acknowledging limitations, such as a small sample size and focus on Iranian SMEs.

Utilization of big data for informed decision-making across various sectors, as highlighted by Jin *et al*., (2020), presents a significant contemporary challenge. While the literature consistently emphasizes the substantial value and competitive advantages organizations can derive from accurate information, there is a potential downside to this increasing dependence on big data. Businesses are actively seeking to leverage big data, aiming to enhance decision-making processes and introduce new technologies for innovative data utilization and knowledge discovery (Storey and Song, 2021). However, the problem lies in the potential overreliance on big data, with organizations possibly neglecting the importance of human judgment and qualitative insights. Relying solely on data-driven decisions may overlook contextual nuances and intangible factors crucial for comprehensive decision-making. Additionally, the challenge extends to ethical considerations, including issues related to data privacy, security, and potential biases within the datasets. The rush to adopt big data solutions without addressing these concerns could lead to unintended consequences and undermine the intended benefits of enhanced decision-making and business value. Therefore, a critical examination of the ethical and human-centric aspects of big data utilization is imperative to ensure responsible and effective implementation across businesses of all sizes.

Sangpetch and Ueasangkomsate's (2023) studied the intricate interplay of big data analytics (BDA), circular economy (CE), and sustainable performance for small and medium-sized enterprises (SMEs). The research uncovers BDA's role as a potent enabler of CE practices within SMEs, utilizing data analysis on material flows, resource consumption, and product lifecycles to inform strategies for

product redesign, waste reduction, and closed-loop systems. The study underscores how embracing CE principles directly enhances SMEs' sustainable performance, reducing environmental impact, improving resource efficiency, and enhancing brand reputation.

The existing literature on Big Data Analytics (BDA) for SMEs' Performance Sustainability in the post-COVID-19 era underscores its transformative potential, emphasizing operational efficiency, market prediction, and long-term sustainability. However, a critical analysis reveals a notable gap in empirical research, particularly in understanding the specific challenges and opportunities faced by small and medium-sized enterprises (SMEs) in the aftermath of the pandemic. The literature lacks comprehensive insights into the practical implementation barriers, adoption challenges, and contextual factors influencing the successful integration of BDA tools within the distinctive operational structures of SMEs navigating the post-COVID-19 business landscape. Additionally, while broad benefits are acknowledged, there is a need for more focused exploration of sector-specific applications and varying impacts across different industries within the SME ecosystem. Closing these gaps would provide valuable insights for SMEs, policymakers, and researchers seeking to foster resilience and sustainability in the evolving post-pandemic business environment

Theoretical Framework

It highlights literature that corresponds with the study's general and specific objectives, fundamental theories correlated to the research and research gaps.

Complex Adaptive System theory provides a perspective for comprehending complex and dynamic systems characterized by numerous interconnected elements, whose behavior can often be unpredictable, as exemplified in various contexts, including businesses. It asserts that these systems are in a continual state of adaptation to their surroundings and are far from static entities. In the context of the study on Data Analytics for SMEs' Performance Sustainability in Post-COVID-19 Kenya, CAS theory gains relevance by recognizing that SMEs function within intricate and swiftly shifting environments. According to Tammisalo (2020), applying CAS theory to SMEs, one gains insights into how data analytics can be harnessed to navigate uncertainties, respond effectively to evolving market dynamics, and uphold sustained performance. This perspective underscores the study's focus on fostering adaptability, agility, and flexibility

in SMEs' adoption of data analytics to bolster their sustainability amid the post-pandemic challenges

The Strategic Choice Theory (SCT), pioneered by Richard Cyert and James G. March postulates that organizational decisions are shaped by a limited pool of information and are heavily influenced by the goals, values, and perceptions of key decision-makers within the organization. SCT underscores the concept of bounded rationality, highlighting that organizations make choices based on cognitive limitations and available information. Hazen, et.al (2020) in their study stated that SCT holds relevance as it sheds light on the decision-making processes of SMEs, particularly concerning adoption of data analytics. It suggests that SME leaders' cognitive processes and perceptions regarding the advantages and risks associated with data analytics will play a pivotal role in determining its adoption. SCT thus informs the study by accentuating the significance of leadership and organizational decision-making in the context of data analytics adoption and its potential influence on SME performance sustainability

Conceptual Framework

This study investigates to what extent do applications of Big Data Analytics, with a focus on Post-COVID-19 circumstances, influence sustainability and performance improvement of small and medium-sized enterprises (SMEs) in Kenya

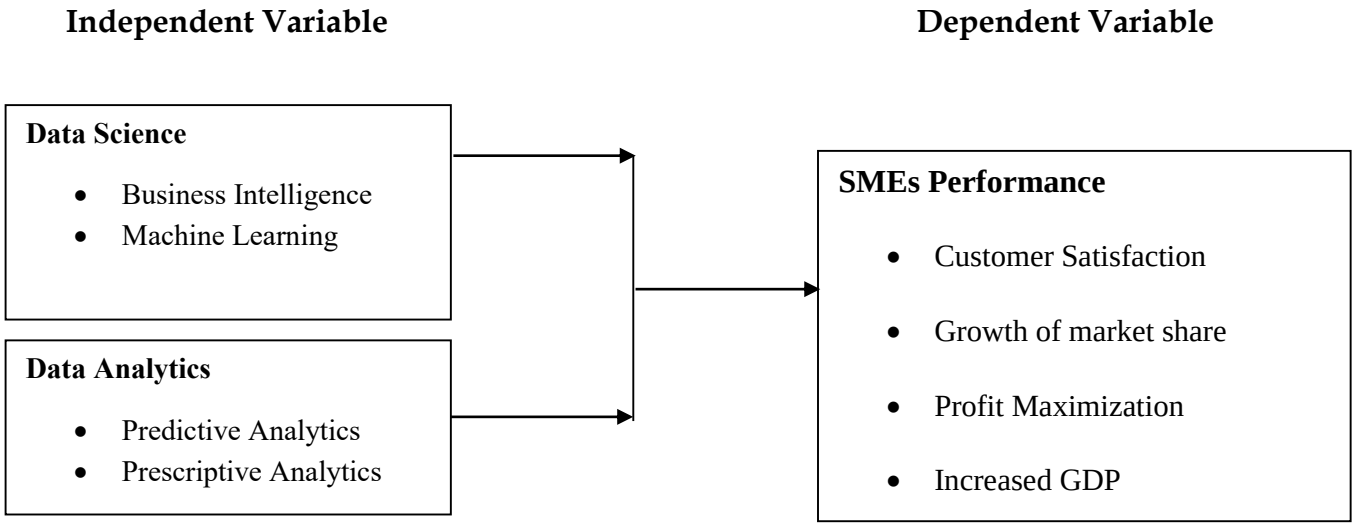

Figure 1: **Conceptual Framework**

Research Methodology

A descriptive survey design was used in this study. The target population was 287 respondents comprising managers of SMEs in Nairobi. The sample for the study was 260 respondents selected using the stratified sampling technique. Questionnaires were used to collect data as shown in Table 3.1.

Table1 Sampling size

Sub-County	Target Population	Sample Size	Percentage (%)
Eastlands	16	14	5.57
Dagoretti North	14	10	4.88
Lang'ata	20	19	6.97
Kibra	17	14	5.92
Roysambu	20	19	6.97
Kasarani	11	10	3.84
Ruaraka	20	19	6.97
Embakasi South	14	14	4.88
Embakasi North		14	6.27
Embakasi Central	18	13	4.53
Embakasi East	13	14	5.92
Embakasi West	17	19	6.62
Makadara	19	10	4.18
Kamuknji	12	19	6.97
Mathare	20	14	5.57
Starehe	16	19	6.97
Dagoretti North	20	19	6.97
Total	287	260	100

Source: Author

Research Findings and Discussions

Table 2 presents data including two key variables, Business Intelligence and Machine Learning, along with their respective mean and standard deviation values. The mean value for Business Intelligence was 3.9 (std. dv = 0.851) indicating that on average, the SMEs in the study exhibited a relatively high level of Business Intelligence utilization in their operations. A mean value above 3 suggests that the majority of SMEs in the sample actively employed Business Intelligence practices to gather insights and make informed decisions. The mean value for Machine Learning is 3.7(std. dv = 0.851) indicating that, on average, SMEs in the study exhibited a reasonably high level of Machine Learning adoption. While the mean is slightly lower than that of Business Intelligence, it still reflects a notable utilization of Machine Learning techniques.

Statement	**Mean**	**Std Dev.**
Business Intelligence	3.9	0.851
Machine Learning	3.7	0.928
Average	**3.8**	**0.8895**

Table 3 below provides descriptive results that focus on the impact of Data Analytics on sustainability of SMEs' performance in Kenya following the COVID-19 pandemic. The table includes two crucial variables, Predictive Analytics and Prescriptive Analytics, along with their respective means and standard deviation values, offering insights into the extent of their utilization among the SMEs in the study. The mean score of 3.73 signifies that the SMEs have a relatively strong adoption of Predictive Analytics. This suggests that a considerable number of SMEs actively used Predictive Analytics to forecast future trends to inform their decision-making and strategic planning. The standard deviation of 0.850 implies moderate variability in the responses regarding Predictive Analytics. While many SMEs employ this practice, there may be some diversity in the extent of its application across the sampled SMEs.

The mean score of 3.85 indicates that the SMEs robustly adopted Prescriptive Analytics to provide specific recommendations and actions based on data analysis. The standard deviation of 0.684 suggests relatively low variability in the responses concerning

Prescriptive Analytics implying that the majority of SMEs exhibited a consistent pattern of adopting Prescriptive Analytics.

Statement	Mean	Std Dev.
Predictive analytics	3.73	0.850
Prescriptive Analytics	3.85	0.684
Average	**3.79**	**0.767**

Table 3: Descriptive Results Data Analytics and Sustainability of SMEs' Performance In Kenya after COVID-19

5.0 Conclusions and Recommendations.

Descriptive results presented on Tables 2 and 3 shed light on the utilization of Data Science and Data Analytics, respectively, in enhancing the sustainability of SMEs' performance in Kenya post-COVID-19. The findings indicate that the SMEs demonstrated notable adoption of Business Intelligence, with a mean score of 3.9, suggesting active utilization to gather insights and make informed decisions. Similarly, the study reveals a reasonably high level of Machine Learning adoption, with a mean score of 3.7. Moving to Data Analytics, the results highlight a strong average adoption of Predictive Analytics (mean of 3.73), indicating widespread use for forecasting future trends, albeit with some variability. These SMEs showcase a robust adoption of Prescriptive Analytics, with an average mean score of 3.85, suggesting consistent utilization for providing specific recommendations and actions based on data analysis. These findings collectively underscore the significance of advanced data-driven techniques in SMEs, portraying a positive landscape for leveraging Data Science and Analytics to enhance sustainability in the post-COVID-19 business environment.

References

Cumba L. T., Huang X., Kholaif M. M. N. H. K. (2020). The impact of digital transformation on big data analytics and firm's sustainability performance in a post-pandemic era. Human Systems Management, (Preprint), 1-22.

Hazen B. T., Skipper J. B., Ezell J. D., Boone C. A. (2020). Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. Computers and Industrial Engineering, 101:592-598.

Jin J., Liu Y., Ji P., Liu H. (2020). Understanding big consumer opinion data for market-driven product design. International Journal of Production Research, 54(10):3019–3041.

KBA Micro, Small and Medium Enterprises (MSMEs) Survey Report, 2021. https://www.centralbank.go.ke/2023/06/13/2022-survey-report-on-msme-access-to-bank-credit/.

Kumar A., Sharma K., Singh H., Naugriya, S. G., Gill, S. S., Buyya, R. (2021). A drone-based networked system and methods for combating coronavirus disease (COVID-19) pandemic. Future Generation Computer Systems, 115, 1-19.

Maroufkhani P., Wan Ismail, W. K., Ghobakhloo M. (2020). Big data analytics adoption model for small and medium enterprises. Journal of Science and Technology Policy Management. 11(4): 483-513.

Nasrollahi, M., Ramezani, J., Sadraei, M. (2021). The impact of big data adoption on SMEs' performance. Big Data and Cognitive Computing, 5(4):68.

Nath, V., Agrawal, R. (2020). Agility and lean practices as antecedents of supply chain social sustainability. International Journal of Operations and Production Management, 40(10):1589-1611.

Sangpetch, P., Ueasangkomsate, P. (2023). The Influence of the Big Data Analytics and Circular Economy on the Sustainable Performance of SMEs. Thammasat Review, 26(1): 114-139.

Storey V. C., Song I.-Y. (2017). Big data technologies and management: What conceptual modeling can do. Data and Knowledge Engineering, 108:50–67.

Tammisalo T. (2020). Harnessing big data for business purposes in Finnish SMEs: adaptive marketing capabilities perspective.

Vitari C., Raguse, E. (2020). Big data analytics business value and firm performance: linking with environmental context. International Journal of Production Research, 58(18): 5456-5476.

AJSTE

Kirinyaga University

AFRICAN JOURNAL OF SCIENCE, TECHNOLOGY AND ENGINEERING (AJSTE)

Published by:

P.O BOX 143 - 10300, KERUGOYA, KENYA

MOBILE +254709742000/+254729499650

Email: info@KyU.ac.ke

Website: www.KyU.ac.ke