Int. j. adv. multidisc. res. stud. 2023; 3(6):3027-3039

Received: 10-11-2023
Accepted: 20-12-2023

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Machine Learning Approaches to Business Process Optimization: Identifying Redundancies and Enhancing Operational Confidentiality in Enterprise IT

[1] Dominic Feboh, [2] Ayokunle Olamide Ijagbemi, [3] Ogochukwu T Izuchukwu, [4] Stanley Nwakamma, [5] Marudi Oyefuga, [6] Nnamdi Chijioke Okeke

[1] Independent Researcher, Calgary, Canada
[2] The Simon Kuznets National University of Economics, Kharkiv, Ukraine
[3] Independent Researcher, Nigeria
[4] Independent Researcher, United States of America
[5] Independent Researcher, Canada
[6] Independent Researcher, Lagos, Nigeria

DOI: https://doi.org/10.62225/2583049X.2023.3.6.6691
Corresponding Author: **Dominic Feboh**

Abstract

This review critically examines the application of machine learning to enterprise business process optimization, with particular emphasis on identifying operational redundancies while preserving confidentiality, privacy, and security. The study adopts a structured narrative review of scholarly literature on business process management, process mining, predictive analytics, artificial intelligence governance, and privacy-preserving computation. The analysis evaluates supervised, unsupervised, deep, reinforcement, and federated learning approaches, alongside anomaly detection, clustering, sequence modelling, natural language processing, and explainable artificial intelligence.

The findings show that machine learning can reveal duplicated tasks, repeated data entry, unnecessary approvals, rework loops, inefficient handoffs, abnormal process variants, and resource-allocation weaknesses. However, repetition alone does not constitute redundancy, because some recurring activities perform essential regulatory, audit, quality-control, or cybersecurity functions. Effective optimization therefore requires contextual interpretation, process-domain expertise, human oversight, and transparent model outputs. The review further establishes that sensitive workflow data can expose proprietary knowledge, employee behaviour, customer information, internal controls, and system vulnerabilities, making confidentiality protection integral to the entire machine-learning lifecycle.

The study concludes that sustainable process improvement depends on the coordinated integration of machine learning, process mining, business process management, cybersecurity, and organisational governance. It recommends phased implementation, reliable data engineering, controlled pilot testing, privacy-by-design, model auditing, multidisciplinary oversight, and continuous performance monitoring. Enterprises should prioritise interpretable, secure, and context-sensitive systems rather than pursue automation as an end in itself. These measures can align efficiency gains with accountability, resilience, compliance, stakeholder trust, and enduring value. Future research should examine causal redundancy detection, privacy-preserving process intelligence, human–machine collaboration, model drift, and implementation conditions across emerging economies and resource-constrained organisational environments.

Keywords: Machine Learning, Business Process Optimization, Process Mining, Operational Redundancy, Confidentiality, Enterprise Governance

1. Introduction

Enterprises operate through networks of activities, platforms, decisions and data exchanges. These arrangements create value but accumulate duplicated approvals, repeated data entry, unnecessary hand-offs, idle time, inconsistent routing and weak controls. Business process management therefore seeks to identify, analyse, redesign and monitor end-to-end work so that organisational objectives are achieved with efficiency, quality and accountability (Dumas *et al.*, 2018) [20]. In enterprise information technology environments, this task is difficult because processes are distributed across enterprise resource planning systems, cloud services and specialised applications. The resulting complexity conceals redundancies within legitimate routines, while fragmented ownership makes it difficult to distinguish necessary controls from avoidable repetition.

Machine learning offers a departure from conventional process improvement, which has often depended on workshops, interviews, static models and manually selected indicators. By learning relationships from historical data, machine-learning models can classify process outcomes, cluster similar cases, detect anomalous paths, estimate remaining time and predict the next activity. Process mining provides a relevant foundation because it extracts behavioural evidence from event logs and connects execution data with process discovery, conformance checking and performance analysis (Van der Aalst, 2016) [66]. Combined with machine learning, such evidence can reveal loops, duplicated sequences, rework and resource-allocation patterns that may not be visible in formal procedures. Predictive monitoring also enables intervention before inefficiencies become completed failures, shifting optimization from retrospective diagnosis to anticipatory control (Teinemaa *et al*., 2019) [63].

The analytical value of this shift is strengthened by deep-learning methods capable of modelling temporal dependencies in process traces. Recurrent neural networks can learn sequential behaviour and predict subsequent events without requiring relationships to be specified in advance (Evermann, Rehse & Fettke, 2017) [22]. This capability is important in enterprise IT, where operational paths vary, exceptions are frequent and cases may cross several systems. Nevertheless, prediction alone does not establish that an activity is redundant. A repeated approval, validation or reconciliation may appear inefficient while serving a regulatory, security or quality-assurance purpose. Effective optimization therefore requires models that combine behavioural regularities with process context, organisational objectives and expert interpretation rather than equating frequency or delay with waste.

Context is important in developing economies, where infrastructure constraints can intensify process fragmentation and resource inefficiency. Nigerian studies of solar integration, hybrid electrical-load distribution and intelligent control strategies illustrate how optimization depends on understanding interdependent resources, control rules and operational constraints rather than improving isolated components (Sunday & Omoegun, 2018; Sunday & Omoegun, 2019; Sunday *et al*., 2019) [55, 56, 57]. Although these studies concern energy and engineering systems, their systems-oriented logic informs enterprise process analysis: local improvements may fail when upstream dependencies, capacity limitations and feedback effects are ignored. For African enterprises, machine-learning adoption must account for variable data quality, infrastructure reliability, skills availability and legacy-system integration. Mhlanga (2021) [38] similarly shows that machine learning can improve data-intensive decision processes while depending on appropriate data and institutional conditions.

Operational confidentiality introduces an important dimension. Enterprise process data may expose customer identities, employee behaviour, financial transactions, commercial strategies, access patterns, supplier relationships and proprietary operating methods. Centralising these records for model training can expand the attack surface and increase the consequences of unauthorised disclosure. Confidentiality risks also arise from trained models, which may retain information about individual records or reveal sensitive patterns through outputs. Differentially private learning addresses part of this problem by limiting the influence of individual training examples and quantifying the privacy cost associated with model development (Abadi *et al*., 2016) [1]. However, privacy protection can reduce model accuracy or complicate implementation, creating a trade-off between analytical utility and information protection that must be governed explicitly.

Federated learning offers another architecture by allowing organisational units or enterprises to train shared models without transferring underlying data to a central repository. This supports data minimisation and may be useful where subsidiaries, departments or partners need collective intelligence but cannot freely exchange sensitive records (Kairouz *et al*., 2021) [33]. Yet decentralisation does not automatically guarantee confidentiality: model updates may leak information, participating systems may be compromised, and heterogeneous data can impair performance. Secure aggregation, access control, encryption, auditability and participation rules are therefore necessary complements. Confidentiality must be designed across the machine-learning lifecycle, including data collection, feature engineering, training, validation, deployment, monitoring and retirement.

1.1 Background and Context of Enterprise Business Process Optimization

Enterprise business process optimization has evolved from periodic workflow redesign into a continuous, data-driven management capability. Contemporary organisations operate through interconnected digital platforms, shared services, cloud applications, automated controls, and distributed teams, creating extensive volumes of operational data. These environments offer considerable opportunities for efficiency, yet they also generate hidden duplication, repeated approvals, fragmented responsibilities, unnecessary handoffs, inconsistent routing, and underused resources. Traditional improvement methods, although valuable, often depend on interviews, static process maps, and retrospective reviews that may not capture how work is actually performed across complex systems. Machine learning strengthens optimization by identifying behavioural patterns, predicting process outcomes, detecting anomalies, and revealing relationships that are difficult to observe manually. When combined with process mining and enterprise information systems, it enables organisations to compare intended workflows with actual execution, locate bottlenecks, and evaluate alternative process configurations. However, optimization is not limited to speed or cost reduction. It also involves maintaining service quality, regulatory compliance, operational resilience, and control. Enterprise process optimization therefore requires an integrated approach that balances intelligent automation with human oversight, secure data use, and organisational accountability.

1.2 Research Problem and Knowledge Gap

Despite growing investment in artificial intelligence and digital transformation, many enterprises struggle with persistent process inefficiencies embedded within routine operations. Redundant tasks are often difficult to identify because they may resemble necessary controls, appear across separate systems, or arise gradually through local adaptations to organisational change. Existing machine learning applications frequently emphasise prediction accuracy, automation potential, or cost savings without adequately explaining whether a detected repetition is wasteful, contextually necessary, or legally required. Another limitation concerns confidentiality. Enterprise process data can contain sensitive information about customers, employees, suppliers, financial transactions, internal controls, system vulnerabilities, and strategic decisions. Yet efficiency-oriented studies often examine such data as an analytical resource while giving insufficient attention to exposure risks during collection, training, deployment, and model interpretation. Privacy-preserving research, by contrast, may focus on technical safeguards without considering their impact on process usefulness, explainability, and implementation. The resulting knowledge gap lies in the limited integration of redundancy detection, performance, confidentiality protection, and governance. A clearer framework is needed to evaluate how machine learning can improve enterprise processes without weakening essential controls or exposing proprietary information.

1.3 Aim, Objectives, and Review Questions

This review aims to examine how machine learning can be applied to optimize enterprise business processes by identifying operational redundancies while preserving the confidentiality of sensitive organisational information. The first objective is to analyse the principal machine learning approaches used in process discovery, prediction, classification, clustering, anomaly detection, and decision support. The second is to evaluate how these methods reveal duplicated activities, repeated data entry, unnecessary approvals, avoidable rework, excessive handoffs, and inefficient resource allocation. The third objective is to assess the confidentiality, privacy, security, and governance risks created when enterprise data are processed for model development. The fourth is to examine privacy-preserving and security-enhancing techniques that support responsible implementation. The review is guided by four questions: Which machine learning techniques are most suitable for enterprise process optimization? How can genuine redundancies be distinguished from necessary control activities? What confidentiality risks arise across the machine learning lifecycle? Which technical and governance mechanisms can protect operational information without undermining analytical value? By addressing these questions, the review seeks to develop an integrated understanding of efficiency, security, explainability, and responsible organisational adoption.

1.4 Scope, Significance, and Structure of the Review

The review focuses on machine learning applications within enterprise information technology environments where operational activities are recorded through event logs, transaction systems, workflow platforms, enterprise applications, communication tools, and digital service infrastructures. Its scope includes supervised, unsupervised, deep, federated, and privacy-preserving learning, together with process mining, predictive monitoring, anomaly detection, natural language processing, and graph-based analysis. Attention is directed to identifying operational redundancies and protecting confidential process information. Broader discussions of artificial intelligence are included only where they directly support these concerns. The review is significant because it connects business process management, data science, cybersecurity, and information governance, fields often examined separately despite their practical interdependence. Its findings are relevant to managers, process analysts, data scientists, enterprise architects, auditors, cybersecurity professionals, policymakers, and researchers seeking secure and accountable optimization. Structurally, the paper first establishes the conceptual foundations of machine learning-driven process improvement. It then examines redundancy detection, performance enhancement, confidentiality protection, implementation challenges, governance requirements, and future research priorities. This arrangement supports an evaluation of how intelligent process optimization can deliver measurable benefits while preserving trust, control, and organisational resilience.

2. Foundations of Machine Learning-Driven Business Process Optimization

Machine learning-driven business process optimization rests on the convergence of business process management, process mining, data engineering and computational learning. Business process management provides the managerial architecture for identifying, modelling, executing, monitoring and redesigning coordinated activities that create organisational value. Its central concern is not automation in isolation, but the disciplined alignment of workflows, resources, controls and information with strategic objectives (Dumas *et al*., 2018) [20]. Machine learning extends this foundation by enabling systems to infer patterns from historical and real-time data, recognise recurring behaviours and support decisions under conditions too complex for fixed rules. Accordingly, optimization becomes a continuous analytical capability rather than an occasional redesign exercise.

At the operational level, enterprise processes generate event data whenever employees, customers, applications or devices interact with digital systems. Typical records contain case identifiers, activities, timestamps, users, resources, outcomes and contextual attributes. Process mining transforms these records into evidence about how work is actually performed, thereby connecting formal process models with observed execution behaviour (Van der Aalst, 2016) [66]. This distinction is fundamental because documented procedures frequently differ from daily practice. Workarounds, repeated approvals, omitted controls, rework loops and informal routing may remain invisible in conventional process maps. Event-log analysis permits the discovery of dominant pathways, deviations, waiting periods and resource dependencies that can subsequently be examined through machine learning.

The analytical foundation includes supervised, unsupervised and reinforcement learning. Supervised methods learn from labelled examples and can classify cases, predict delays, estimate completion times or identify transactions likely to require rework. Unsupervised methods search for structure without predetermined labels, making clustering and anomaly detection valuable for identifying unusual process variants, duplicated task patterns and previously unknown inefficiencies. Reinforcement learning focuses on sequential decision-making, where an agent learns policies by evaluating the consequences of actions over time. These approaches are supported by algorithms such as decision trees, random forests, support vector machines, neural networks and association methods, each offering different balances of accuracy, interpretability, data requirements and computational cost (Sarker, 2021) [52].

Deep learning is particularly relevant where enterprise processes contain long, nonlinear and highly variable sequences. Recurrent neural networks can learn temporal dependencies among events and predict the next activity, responsible resource or likely continuation of an active case (Evermann, Rehse & Fettke, 2017) [22]. Predictive process monitoring broadens this capability by estimating future outcomes while cases are still running, allowing organisations to intervene before delays, compliance failures or service breakdowns become irreversible (Teinemaa *et al*., 2019) [63]. Such models can therefore move process management from retrospective reporting toward anticipatory coordination. Nevertheless, highly accurate predictions are not automatically operationally useful. Outputs must be timely, interpretable and connected to decisions that process owners are authorised and equipped to make.

Process heterogeneity is another important foundation. The same nominal workflow may be executed differently across departments, countries, customer groups or technology platforms. Process variant analysis compares these differing executions to determine which pathways are faster, more reliable or more compliant, and to explain why performance varies (Taymouri *et al*., 2021) [62]. This is crucial for redundancy identification because repetition cannot be judged solely by frequency. A duplicated activity may reflect poor system integration, but it may also represent a legitimate segregation-of-duties control, regulatory checkpoint or risk response. Machine learning must therefore be combined with contextual variables and domain expertise to distinguish non-value-adding duplication from necessary operational safeguards.

The predictive maintenance domain illustrates the broader logic of machine learning-driven optimization. Instead of waiting for equipment failure or servicing assets at fixed intervals, predictive systems analyse condition data to forecast deterioration and schedule intervention according to actual need. Nigerian research demonstrates how artificial intelligence, sensor information and data analytics can shift maintenance from reactive response to proactive decision-making, reducing downtime and improving resource utilisation (Sunday *et al*., 2020) [58]. Related work on thermodynamic efficiency and control strategies shows that optimized performance depends on coordinated monitoring, feedback and adaptive control rather than isolated technical adjustments (Sunday *et al*., 2019) [57]. Although these studies address engineering systems, their principles apply to enterprise workflows: effective optimization requires continuous observation, reliable data, early detection and context-sensitive intervention.

African applications also demonstrate that machine learning must be situated within institutional and infrastructural realities. In South Africa, machine learning has been examined as a means of improving credit-risk assessment by analysing alternative data and reducing information asymmetry in emerging-market finance (Mhlanga, 2021) [38]. The example highlights the capacity of machine learning to enhance complex, data-intensive decisions, but also reveals dependence on data quality, regulatory acceptability and organisational competence. Similar considerations apply across enterprise IT, particularly where fragmented legacy systems, limited interoperability and uneven digital maturity constrain model development.

2.1 Enterprise Business Processes, Workflow Data, and Operational Redundancy

Enterprise business processes are coordinated sequences of activities through which organisations convert information, resources and decisions into products, services and regulatory outcomes. In enterprise IT, these processes rarely reside within a single application. They traverse enterprise resource planning platforms, customer relationship management systems, workflow engines, cloud services and databases. Each interaction leaves digital evidence that can be reconstructed as workflow data. Properly interpreted, these records reveal what the organisation intends to do, but how work is actually executed and which activities consume resources without improving the final outcome.

Workflow data consist of case identifiers, activity labels, timestamps, users, transaction attributes, documents, costs and process outcomes. Their analytical value depends on completeness, consistency and event-to-case correlation. Missing timestamps, duplicated records, inconsistent naming conventions, activity labels and events captured at different levels of granularity can distort process representations. Marin-Castro and Tello-Leal (2021) [36] show that preprocessing is essential before event logs are used for discovery or conformance analysis. Cleaning, harmonisation, abstraction and trace construction are technical preparations; they determine whether machine-learning models recognise genuine operational behaviour or learn artefacts created by defective data.

Process discovery converts event logs into representations of observed workflow behaviour. Automated discovery methods infer task ordering, branching, concurrency, loops and alternative pathways, thereby exposing differences between prescribed procedures and operational reality. However, logs often generate complex models because processes contain exceptional cases, inconsistent labels and execution patterns. Augusto *et al*. (2019) [9] demonstrate that discovery methods must balance fitness, precision, simplicity and generalisation. A model that reproduces every recorded variation may become unintelligible, while an excessively simplified model may conceal important controls. This balance is central when the purpose is to identify redundancy rather than merely visualise process flow.

Operational redundancy refers to work that is repeated, duplicated or unnecessarily prolonged without delivering proportionate value. It may appear as repeated data entry across disconnected systems, duplicate customer records, unnecessary approvals, overlapping responsibilities, repeated document verification, excessive handoffs, avoidable rework or parallel applications performing equivalent functions. Redundancy can also be semantic: two activity labels may describe the same action because departments use different terminology. Such inconsistencies inflate event-log complexity and may cause algorithms to interpret equivalent tasks as distinct operations. Anomaly-detection methods can identify infrequent or structurally unusual traces, although rarity alone does not establish inefficiency (Bezerra & Wainer, 2013) [12]. Exceptional pathways may be justified by risk or regulation.

Machine learning strengthens redundancy analysis by comparing traces, clustering similar process variants and detecting recurrent combinations associated with delay, error or rework. Predictive models can estimate whether an active case is likely to repeat an activity, miss a deadline or require escalation. Long short-term memory networks, for example, can learn sequential dependencies from completed cases and predict subsequent events and remaining process duration (Tax *et al*., 2017) [61]. These capabilities allow process owners to intervene before redundant behaviour is completed. Nevertheless, prediction must be interpreted alongside process purpose. A repeated control may protect confidential information, enforce segregation of duties or provide evidence for audit, making its apparent duplication operationally necessary.

Robotic process automation offers a complementary response by reproducing structured user actions across existing applications. It is particularly applicable to high-volume, rule-based activities such as copying data, reconciling fields, generating routine reports and transferring information between systems. Syed *et al*. (2020) [60] observe that automation can relieve employees of repetitive work, but poorly selected processes may simply automate existing inefficiency. Consequently, organisations should first determine whether an activity should be removed, redesigned, integrated or automated. Machine learning can support this decision by measuring frequency, variation, error incidence and downstream effects rather than treating repetition as sufficient evidence for automation.

Evidence from African organisations reinforces the importance of contextual diagnosis. Greyling and Jooste (2017) [27] demonstrate in South Africa that process mining can reveal execution patterns and improvement opportunities within physical asset management. Nkomo and Marnewick (2021) [41] further emphasise that process redesign requires clearly defined roles, methods, communication and change-management arrangements. In Nigeria, Agbionu and Audu (2021) [3] associate work-process innovation with organisational performance in manufacturing firms, supporting the practical value of periodically examining established routines. These perspectives are significant where legacy systems, infrastructure constraints and fragmented responsibilities may generate duplication that cannot be resolved through algorithmic analysis alone.

Workflow data also contain confidentiality risks. Event logs may disclose customer identities, employee behaviour, internal controls, commercially sensitive transactions and security-related actions. Even when direct identifiers are removed, combinations of activities, timestamps and attributes can permit re-identification. Nuñez von Voigt *et al*. (2020) [42] show that individual process traces may possess substantial uniqueness, making conventional anonymisation insufficient. Redundancy analysis must therefore apply data minimisation, role-based access, secure preprocessing and controlled reporting. The enterprise should preserve enough detail to distinguish waste from necessary control while limiting exposure of sensitive operational relationships. Thus, workflow intelligence depends on disciplined data engineering, contextual interpretation and confidentiality-aware governance as much as on machine-learning performance.

2.2 Machine Learning Paradigms and Techniques for Process Analysis

Machine learning paradigms provide the mechanisms through which enterprise workflow data become evidence about process behaviour, inefficiency and risk. Their value lies in learning relationships among events, resources, timing, case attributes and outcomes without requiring every operational rule to be programmed manually. Machine learning should therefore be understood as complementary approaches rather than a single optimization instrument. Jordan and Mitchell (2015) [32] distinguish learning systems by the experience available and the task pursued, while Marwala (2018) [37] stresses that algorithm selection must reflect the problem, data and organizational context.

Supervised learning is suitable when historical process cases possess known labels or measurable outcomes. Classification models can predict whether a transaction will be delayed, rejected, escalated or returned for rework, whereas regression models can estimate cycle time, cost or resource consumption. Decision trees are comparatively transparent, but individual trees may be unstable when workflow data are noisy. Random forests mitigate this weakness by aggregating randomized trees, improving robustness and accommodating nonlinear relationships among process variables (Breiman, 2001) [16]. Within enterprise workflows, such models can identify attributes associated with repeated approvals, service failure or excessive handling time. Their usefulness depends on representative labels, since inconsistent historical decisions may reproduce organizational deficiencies.

Unsupervised learning is valuable where reliable labels are absent or where enterprises seek unknown patterns. Clustering can group process traces according to activity sequences, duration, resource use or outcome similarity, revealing dominant variants and unusual pathways. Anomaly detection is relevant to redundancy analysis because it can flag loops, repeated transactions, abnormal handoffs and rare activity combinations. Isolation Forest identifies anomalies by recursively partitioning observations and exploiting the tendency of unusual cases to be isolated more rapidly than normal ones (Liu, Ting & Zhou, 2008) [35]. However, an anomalous trace is not automatically defective; exceptional cases may represent legitimate responses to complex customers, compliance requirements or security incidents.

Semi-supervised learning combines a limited labelled dataset with a larger body of unlabelled process records. It is useful where expert annotation is expensive, confidential or disruptive. Active learning reduces annotation burdens by selecting uncertain cases for expert review. These paradigms support model development for rare failures, suspicious behaviour or specialised exceptions. Process owners remain necessary to distinguish avoidable repetition from necessary control.

Deep learning extends process analysis by learning representations directly from sequential and high-dimensional data. Long short-term memory networks preserve information across extended sequences and address limitations in learning long-range dependencies (Hochreiter & Schmidhuber, 1997) [29]. Applied to event logs, they can model activity order and timing, predict the next event, estimate remaining duration and detect sequences likely to produce undesirable outcomes. Natural language processing can complement sequence models by analysing service tickets, emails, policies and task descriptions to identify semantically duplicated work. Their principal limitation is opacity: highly parameterised models may generate accurate predictions without giving managers a clear account of the operational reasoning involved.

Reinforcement learning shifts attention from prediction to sequential intervention. An agent evaluates actions through rewards and learns policies intended to improve cumulative performance. Deep reinforcement learning demonstrates that complex control policies can be learned from high-dimensional observations by combining representation learning with reward-driven adaptation (Mnih *et al*., 2015) [40]. In enterprise processes, possible actions include reallocating resources, reprioritising cases, bypassing avoidable steps or escalating high-risk instances. Rewards must be designed carefully. A policy focused narrowly on speed may remove confidentiality checks, overload employees or increase downstream errors. Safe application therefore requires constraints reflecting compliance, service quality, fairness and information security.

Hybrid techniques often provide greater practical value than isolated paradigms. Predictive maintenance illustrates this integration because sensor data, classification, anomaly detection and forecasting can anticipate failure and schedule intervention. Nigerian research on the transition from reactive to predictive maintenance demonstrates how data-driven monitoring can reduce downtime and improve technical resource use (Sunday *et al*., 2020) [58]. The same logic applies to enterprise workflows, where patterns preceding rework, bottlenecks or system failure can be detected early enough for corrective action. Models must also be updated as workflows and user behaviour change, because concept drift can weaken predictions.

Interpretability connects analytical performance with accountable process management. Local explanation methods can approximate complex predictions around individual cases and show which variables influenced an output (Ribeiro, Singh & Guestrin, 2016) [49]. Such explanations help managers determine whether a model identified genuine redundancy or relied on sensitive or misleading attributes. Technique selection should therefore consider accuracy, interpretability, scalability, confidentiality, data requirements and intervention costs together, providing a balanced basis for integrating machine learning with enterprise process analysis. This balance is essential for defensible, secure and context-sensitive sustainable enterprise process optimization.

2.3 Integration of Process Mining, Business Process Management, and Machine Learning

The integration of process mining, business process management and machine learning creates a unified analytical architecture for understanding, controlling and improving enterprise operations. Business process management supplies the lifecycle through which processes are identified, modelled, executed, monitored and redesigned, whereas process mining grounds that lifecycle in event data generated by information systems. Machine learning extends both disciplines by recognising patterns, predicting outcomes and recommending interventions. This integration is valuable because formal process documentation often represents intended behaviour, while event logs reveal actual execution, including workarounds, delays, repeated activities and unauthorised deviations. The Process Mining Manifesto positions process mining between computational intelligence and process management, emphasising its capacity to connect recorded events with operational models (van der Aalst *et al*., 2012) [67].

At the discovery stage, process mining converts event logs into models that describe activity sequences, choices, parallelism and loops. Block-structured discovery methods can produce comprehensible models that remain suitable for subsequent analysis and redesign (Leemans, Fahland & van der Aalst, 2013) [34]. Machine learning contributes by clustering similar traces, detecting unusual variants and learning relationships between process characteristics and performance outcomes. The resulting evidence enables business process managers to determine whether repeated activities represent genuine redundancy, exceptional but legitimate handling, or weaknesses in system integration. This distinction prevents organisations from removing controls merely because they occur frequently or increase processing time.

Conformance checking provides the second integration point. It compares observed executions with prescribed process models to identify deviations involving activity order, resource allocation, timing or data conditions. Multi-perspective alignment techniques allow operational traces to be evaluated against control-flow and contextual requirements rather than against sequence alone (De Leoni & van der Aalst, 2013) [18]. Machine-learning models can then examine recurring deviations and estimate their association with rework, service failure, confidentiality breaches or excessive cost. Business process management translates these findings into governance responses, including redesign, policy revision, staff training and automated controls. Thus, deviations become evidence for managed improvement rather than isolated technical exceptions.

Integration is equally important across organisational boundaries. Enterprise processes frequently depend on suppliers, subsidiaries, customers and external service providers whose activities are recorded in separate systems. Collaborative workflow mining can combine multiple event sources to reconstruct cross-organisational processes and expose handoff delays, duplicated validation and inconsistent information exchange (Zeng *et al*., 2013) [71]. Machine learning can identify patterns across these distributed traces, while BPM establishes ownership and accountability for changes. However, data integration must be governed carefully because combined logs may reveal commercially sensitive relationships, individual behaviour or internal control structures.

Predictive process monitoring further shifts BPM from retrospective assessment towards proactive management. By analysing incomplete traces, machine-learning models can estimate remaining time, likely outcomes, next activities and the probability of undesirable events. A value-driven selection framework is necessary because different prediction methods suit different data structures, business questions and intervention opportunities (Di Francescomarino *et al*., 2018) [19]. Predictions become operationally useful only when incorporated into process rules, dashboards, escalation procedures and decision rights. Integration therefore requires more than model deployment;

it requires a clear mechanism linking analytical signals to authorised organisational action.

Business intelligence strengthens this architecture by presenting process indicators, predictions and comparative performance information to managers. Evidence from European small and medium-sized enterprises indicates that alignment between BPM and business intelligence is associated with stronger organisational performance than isolated implementation (Pejić Bach *et al*., 2019) [45]. Similar implications arise in Nigeria, where process redesign has been associated with improved operational performance in food and beverage firms (Olajide & Okunbanjo, 2020) [44]. These findings underline that analytical technologies create value when embedded within process ownership, organisational resources and continuous improvement routines.

African research also demonstrates the need for adaptive integration rather than technology-centred redesign. Ahmed Bayomy, Khedr and Abd-Elmegid (2021) [4] propose an adaptive model that diagnoses breakdowns in business process reengineering and supports context-sensitive corrective action. For enterprise IT, this reinforces the importance of combining event-log evidence, predictive models, managerial judgement and feedback mechanisms. Effective integration should maintain traceability from raw data to discovered models, predictions, interventions and measured outcomes. It should also incorporate access controls, data minimisation, validation and human oversight so that optimisation does not compromise operational confidentiality. Through this coordinated architecture, process mining explains execution, machine learning anticipates performance, and BPM governs responsible organisational change over time.

3. Machine Learning Approaches for Identifying Process Redundancies

Machine learning approaches for identifying process redundancies operate by converting enterprise event data into evidence about duplicated, unnecessary or inefficient work. Redundancy may occur when equivalent activities carry different labels, the same task is repeated across systems, approvals are duplicated, cases circulate through rework loops, or resources perform overlapping responsibilities. Detecting these conditions requires more than counting repeated events. A frequent activity may be essential to compliance or confidentiality, whereas a rare pathway may expose avoidable escalation. Machine learning must therefore evaluate behavioural similarity, sequence context, semantic meaning and performance consequences together.

Clustering is particularly useful when organisations possess heterogeneous process variants without reliable labels. Trace-clustering methods group cases according to activity order, timing, resources or contextual attributes, allowing analysts to separate dominant workflows from specialised or inefficient variants. Context-aware clustering improves process discovery by partitioning dissimilar traces before models are generated, reducing the "spaghetti" effect created when highly variable behaviour is represented in one model (Bose & van der Aalst, 2009) [15]. Within redundancy analysis, clusters can reveal teams that perform equivalent work through different routes, cases with excessive handoffs, and recurring sequences associated with delay or rework. However, clustering remains exploratory; domain experts must determine whether the identified variation is unnecessary or operationally justified.

Activity-label analysis addresses semantic redundancy. Enterprise logs assembled from multiple applications frequently contain synonymous labels, abbreviations and inconsistent naming conventions. These differences can make one operational action appear to be several separate activities. Duplicate-task detection techniques examine the local behavioural context surrounding repeated labels and can refine discovered models by splitting or consolidating tasks where appropriate (Vázquez-Barreiros, Mucientes & Lama, 2016) [68]. Natural language processing can extend this logic to free-text descriptions, service tickets, policy documents and approval narratives, using lexical or semantic similarity to identify tasks that express the same function despite different wording.

Anomaly detection and filtering methods provide another analytical route. Infrequent behaviour can introduce model complexity and obscure dominant process structures. Automated filtering can remove behaviour that is statistically exceptional while retaining patterns necessary for representative process discovery (Conforti, La Rosa & ter Hofstede, 2017) [17]. Yet indiscriminate filtering may eliminate genuine fraud indicators, critical exceptions or confidentiality controls. Event-log quality is therefore decisive. Missing events, incorrect timestamps, polluted labels and incorrectly correlated cases can create false evidence of duplication or conceal real redundancy (Suriadi *et al*., 2017) [59]. Machine-learning pipelines should incorporate data-quality assessment, provenance checks and sensitivity analysis before recommendations are presented.

Supervised learning becomes valuable when organisations can label previous activities as necessary, redundant, erroneous or rework-related. Classification models may learn combinations of duration, resource usage, sequence position and outcome that distinguish useful controls from non-value-adding repetition. Predictive models can then flag active cases likely to repeat an activity, exceed an approval threshold or return to an earlier stage. Al-Anqoudi *et al*. (2021) [5] argue that machine learning can reduce the cost and complexity of business process re-engineering by exploiting patterns within process data. Nevertheless, historical labels may encode inconsistent managerial decisions, so model validation should include process owners, compliance specialists and information-security personnel.

The practical value of redundancy identification is illustrated by digital process redesign. Garcia-Garcia *et al*. (2021) [26] demonstrate how re-engineering and digitalising quality-control checks can reduce physical waste and resource consumption while improving the visibility of operational activities. The lesson for enterprise IT is that redundancy detection should lead to a clearly designed response: removing an activity, integrating systems, standardising terminology, reallocating responsibility or automating a retained task. Automation should not preserve an unnecessary process merely because it can be executed more quickly.

Organisational context remains especially important in African enterprises. Evidence from Nigerian airline companies indicates that business process re-engineering is materially connected with operational cost structures, reinforcing the importance of examining inefficient routines rather than treating technology as an isolated solution

(Adefulu *et al*., 2020) [2]. South African research similarly shows that digital transformation requires leadership, suitable organisational structures, technical capability and coherent implementation strategy (Zulu, Pretorius & van der Lingen, 2021) [72]. Accordingly, machine-learning findings should be interpreted within infrastructure, workforce, regulatory and governance conditions. A defensible redundancy-identification framework combines statistical evidence with process purpose, confidentiality requirements, causal consequences and human review, ensuring that efficiency gains do not remove controls that sustain security, accountability or service quality overall.

4. Machine Learning for Process Enhancement and Operational Performance

Machine learning enhances enterprise processes by converting historical and real-time operational data into predictions, recommendations and adaptive controls. Its contribution extends beyond automating repetitive tasks because models can forecast demand, prioritise cases, estimate completion times, allocate resources and detect conditions likely to produce delay, rework or service failure. Process enhancement becomes most effective when analytical outputs are connected to measurable objectives such as cycle time, throughput, cost, accuracy, customer responsiveness, compliance and resource utilisation. A process-oriented organisation is better positioned to translate these outputs into performance gains because responsibilities, information flows and improvement priorities are defined across functional boundaries (van Assen, 2018) [65].

Predictive process monitoring supports early intervention while work is still underway. Models trained on event logs can estimate an active case's likely outcome, remaining duration or next activity, enabling managers to redirect resources before inefficiency becomes irreversible. Prescriptive monitoring advances this capability by evaluating alternative actions against operational key performance indicators. Weinzierl *et al*. (2020) [70] demonstrate that next-best-action recommendations can be designed around performance objectives while preserving process-flow requirements. This approach is particularly valuable for congested service operations, where timely reprioritisation or reassignment can prevent cascading delays.

The performance value of machine learning nevertheless depends on organisational capability rather than algorithmic accuracy alone. Mikalef and Gupta (2021) [39] conceptualise artificial intelligence capability as a combination of data, technological, human and organisational resources that supports creativity and firm performance. Likewise, Jöhnk, Weißert and Wyrtki (2021) [31] identify data availability, skills, strategic alignment, governance and change readiness as critical conditions for effective artificial intelligence adoption. These insights indicate that enterprises require reliable infrastructure, competent personnel and accountable decision rights before predictions can be converted into sustained process improvement.

Machine-learning transformation can improve administrative, operational and customer-facing activities when projects are aligned with business value. Wamba-Taguimdje *et al*. (2020) [69] show that artificial intelligence-based transformation can strengthen firm performance by reconfiguring process capabilities rather than merely installing new technology. This distinction is important because automating a poorly designed workflow may accelerate waste, duplicate errors or embed unsuitable decisions. Enhancement should therefore begin with process diagnosis, followed by selective automation, model integration and continuous performance assessment.

Explainability is also essential to operational performance. Managers need to understand why a model predicts delay, recommends reassignment or identifies a case as high risk. Galanti *et al*. (2020) [25] show how predictive process monitoring can be supplemented with explanations that identify influential process features. Transparent outputs support informed intervention, model validation and employee acceptance, while helping organisations detect reliance on misleading or confidential attributes. Human review remains necessary where recommendations affect regulatory obligations, employee workloads or access to sensitive enterprise information.

Evidence from Nigeria reinforces the relationship between process redesign and organisational results. Eze, Adelekan, and Nwaba (2019) [23] report that components of business process reengineering positively influenced the performance of Nigerian insurance firms. Although process reengineering is broader than machine learning, the finding demonstrates that technology produces value when combined with redesigned routines and managerial commitment. Accordingly, operational performance should be evaluated through balanced measures, including efficiency, quality, resilience, confidentiality and stakeholder outcomes. Machine learning is therefore most effective as a governed decision-support capability that continuously learns from process execution while preserving human accountability and organisational control. Regular model recalibration further ensures that recommendations remain relevant as process conditions, demand patterns and organisational priorities evolve.

5. Operational Confidentiality, Privacy, and Security in Machine Learning Systems

Operational confidentiality in machine-learning systems concerns the protection of enterprise information throughout data collection, preparation, model training, deployment, monitoring and retirement. Business process optimization commonly relies on event logs, transaction histories, employee actions, customer records, system configurations and internal control data. These sources can reveal proprietary workflows, commercial strategies, access privileges and organisational vulnerabilities. Consequently, a model that improves cycle time or detects redundancy may simultaneously enlarge the exposure surface if sensitive records are centralised, copied across environments or accessed by inadequately authorised personnel.

Privacy risk persists even when direct identifiers are removed. Machine-learning models can memorise characteristics of their training data, permitting attackers to determine whether a particular record was used during training. Membership inference attacks demonstrate that prediction outputs may disclose participation in a dataset, especially where models are overfitted or provide detailed confidence scores (Shokri *et al*., 2017) [53]. Model inversion presents a related danger because adversaries may use model responses to reconstruct sensitive attributes or representative inputs (Fredrikson, Jha & Ristenpart, 2015) [24]. In enterprise process analysis, these attacks could expose employee behaviour, customer status, operational exceptions or confidential control patterns.

Security risks also threaten model integrity and availability. Poisoning attacks manipulate training data or updates, while evasion attacks alter inputs to produce misleading predictions. Such interference may cause a process-optimization system to classify necessary controls as redundant, overlook fraudulent activity or recommend unsafe resource allocations. Biggio and Roli (2018) [13] therefore emphasise adversarial threat modelling and security-by-design evaluation rather than assuming that high predictive accuracy implies operational trustworthiness. Enterprise controls should include dataset provenance, input validation, model testing, access logging, version management and continuous monitoring for abnormal behaviour.

Privacy-preserving techniques can reduce exposure while retaining analytical value. Federated learning allows organisational units to train a shared model without transferring raw data to a central repository. Nevertheless, model updates can leak information or be poisoned by compromised participants, requiring secure aggregation, authentication and robust update validation (Blanco-Justicia *et al*., 2021) [14]. Differential privacy, encryption, pseudonymisation, secure multiparty computation and data minimisation provide additional safeguards, although each introduces trade-offs involving accuracy, computational cost, interpretability and implementation complexity.

Governance must translate these technical protections into accountable practice. Nigerian scholarship highlights weaknesses that can arise when data-protection obligations, enforcement mechanisms and organisational awareness remain insufficiently developed (Odusote, 2021) [43]. Across Africa, artificial-intelligence deployment raises broader concerns regarding policy capacity, inclusion, infrastructure and privacy governance (Gwagwa *et al*., 2020) [28]. South African evidence further shows that data exposure is a prominent cyber incident affecting public and private organisations, reinforcing the need for practical prevention and response capability (Pieterse, 2021) [46].

6. Implementation Challenges, Governance, and Organisational Readiness

Implementing machine learning for business process optimization is rarely constrained by algorithm selection alone. The principal challenge is converting experimental models into dependable organisational capabilities that operate across fragmented systems, changing workflows and sensitive data environments. Enterprise adoption therefore requires coordinated attention to technology, data, people, governance and institutional context. The technology–organisation–environment perspective is useful because adoption depends simultaneously on technical suitability, internal resources and external pressures rather than on perceived usefulness alone (Awa, Ojiabo & Orokor, 2017) [10]. In South African small and medium-sized enterprises, infrastructure, managerial support, skills and environmental conditions similarly shape technology adoption readiness (Jere & Ngidi, 2020) [30].

Technical implementation begins with data quality and systems integration. Process data are often distributed across enterprise applications, recorded at inconsistent levels of detail and affected by missing values, duplicated events or unstable labels. Sambasivan *et al*. (2021) [51] describe how unresolved data problems can propagate through artificial intelligence projects as "data cascades", producing downstream failures that remain difficult to diagnose. Legacy platforms create additional difficulties because models must interact with existing databases, workflow engines and security controls without disrupting operational continuity. Amershi *et al*. (2019) [6] show that machine-learning development introduces software-engineering challenges involving data dependencies, testing, monitoring, versioning and collaboration between specialised roles. These requirements make deployment a lifecycle discipline rather than a one-time technical installation.

Organisational readiness also depends on leadership commitment, workforce competence and realistic expectations. Process owners must understand where predictions or recommendations enter the workflow, who may override them and how outcomes will be evaluated. Employees may resist systems perceived as threatening professional autonomy, increasing surveillance or redistributing responsibilities without consultation. Training must therefore extend beyond technical teams to managers, operational staff, auditors and security personnel. Berente *et al*. (2021) [11] argue that managing artificial intelligence requires attention to its autonomy, learning capacity and inscrutability. These characteristics create continuing managerial obligations because models may change, generate unexpected behaviour or become difficult to explain after deployment.

Decision rights constitute another implementation challenge. Machine learning can automate routine classifications, augment human judgement or restructure how authority is distributed. The appropriate arrangement depends on decision complexity, interpretability, speed, replicability and the consequences of error (Shrestha, Ben-Menahem & von Krogh, 2019) [54]. High-impact process changes, particularly those affecting confidentiality, compliance, employees or customers, should retain meaningful human review. Clear escalation routes are required when model outputs conflict with policy, professional judgement or emerging operational evidence.

Governance provides the mechanisms through which these responsibilities are assigned and enforced. Effective arrangements should define data ownership, model ownership, approval authority, documentation standards, access controls, validation procedures and performance thresholds. Internal algorithmic auditing can embed accountability across design, development and deployment by requiring evidence of objectives, assumptions, testing decisions and identified risks (Raji *et al*., 2020) [48]. Governance should also require monitoring for accuracy deterioration, concept drift, bias, security incidents and unintended process consequences. Models that no longer satisfy operational or confidentiality requirements must be recalibrated, restricted or retired.

7. Future Research Directions and Implications for Enterprise Practice

Future research on machine learning-driven business process optimization should move beyond demonstrations of predictive accuracy towards evidence of sustained organisational value. Studies evaluate models on static datasets, yet enterprise processes evolve as regulations, technologies, workloads and employee practices change. Longitudinal research is therefore required to examine concept drift, model deterioration, unintended workarounds and the durability of efficiency gains. Comparative studies across industries and institutional settings should also determine whether redundancy-detection methods developed in data-rich organisations remain reliable in smaller firms, public institutions and emerging economies. Digital transformation research indicates that successful change depends on organisational capabilities, structures and strategic coordination, not merely technology acquisition (Verhoef *et al*., 2021).

A major research priority concerns the distinction between harmful redundancy and necessary operational control. Models should incorporate causal inference, process semantics, and domain constraints so that repeated approvals, validations, or handoffs are not removed solely because they increase cycle time. Research should test whether proposed interventions actually cause improvements rather than merely correlate with high-performing process traces. In high-stakes settings, inherently interpretable models may be preferable to post-hoc explanations because managers must understand and defend recommendations affecting compliance, confidentiality and resource allocation (Rudin, 2019) [50]. Researchers should consequently develop evaluation frameworks that measure interpretability, causal validity, fairness, security and operational usefulness alongside conventional predictive metrics.

Privacy-preserving process intelligence represents another critical direction. Cross-organisational learning could allow enterprises, subsidiaries and supply-chain partners to improve shared processes without centralising confidential records. However, future work must evaluate whether federated, anonymised or encrypted learning preserves sufficient information to detect meaningful redundancy. Studies should measure the trade-offs among confidentiality, data utility, computational cost and process-model fidelity. Research is also needed on governance mechanisms for model updates, participant verification, auditability and liability when distributed systems generate harmful recommendations. These questions reflect wider multidisciplinary concerns that artificial intelligence research must integrate technical, ethical, managerial and policy perspectives (Dwivedi *et al*., 2021) [21].

Human–machine collaboration should receive equal attention. Artificial intelligence may automate repetitive activities while augmenting judgement in ambiguous cases, but these functions are interdependent and can create organisational tension (Raisch & Krakowski, 2021) [47]. Studies should examine how employees contest, override, learn from and adapt to machine-generated process recommendations. Experimental and field research should identify suitable levels of autonomy, effective explanation formats and safeguards against automation bias. Enterprises should retain accountable process owners and establish escalation pathways for cases involving confidential information, regulatory obligations or substantial stakeholder consequences.

The African context requires locally grounded investigation rather than the direct transfer of models designed for technologically mature environments. Research on artificial intelligence in Africa highlights institutional capacity, digital inequality and context-sensitive policy as decisive considerations (Arakpogun *et al*., 2021) [7]. Nigerian evidence similarly indicates that digital transformation outcomes depend on stakeholder commitment, multidisciplinary coordination and effective governance (Ufua *et al*., 2021) [64]. African enterprise studies should examine infrastructure reliability, informal work practices, multilingual data, limited labelled datasets and scarce specialist expertise. Atiku and Obagbuwa (2021) [8] demonstrate the feasibility of machine-learning classification for organisational performance analysis in African banking contexts, but broader validation is required across sectors and countries.

8. Conclusion

This review examined how machine learning can be deployed to improve enterprise processes by identifying redundant activities, strengthening operational performance, and preserving the confidentiality of sensitive organisational information. The study's aims and objectives were achieved through a structured synthesis of business process management, process mining, predictive analytics, privacy engineering, and governance perspectives. The analysis established that machine learning can detect duplicated tasks, recurrent rework, unnecessary approvals, inefficient handoffs, abnormal process variants, and resource-allocation weaknesses when reliable workflow data are available. It also showed that supervised, unsupervised, deep, reinforcement, and privacy-preserving learning techniques contribute differently to prediction, anomaly detection, sequence analysis, and process redesign.

A central finding is that redundancy cannot be judged by repetition alone. Activities that appear inefficient may perform essential regulatory, security, audit, or quality-control functions. Accordingly, machine-generated recommendations require contextual interpretation, process-domain expertise, explainability, and human oversight. The review further demonstrated that operational confidentiality must be embedded throughout the machine-learning lifecycle. Event logs, transaction records, employee actions, and process models can expose proprietary knowledge, making access control, data minimisation, encryption, federated learning, differential privacy, secure aggregation, and continuous security monitoring indispensable.

The study also found that successful adoption depends on organisational readiness, high-quality data, interoperable infrastructure, clear decision rights, multidisciplinary collaboration, and robust governance. Enterprises should therefore avoid technology-led implementation that merely automates existing inefficiencies. Instead, they should adopt phased, problem-driven programmes supported by baseline metrics, controlled pilots, model auditing, employee engagement, and clearly defined escalation procedures. Future implementation should prioritise interpretable and context-sensitive models, especially in high-stakes environments and emerging economies. Overall, machine learning offers substantial potential to create faster, more resilient, secure, and evidence-based enterprise operations, provided that optimisation objectives remain aligned with confidentiality, accountability, compliance, and sustained organisational value. This balanced approach provides a foundation for responsible innovation and enduring advantage across increasingly complex digital enterprises.

9. References

1. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, *et al*. Deep learning with

differential privacy', In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, 308-318. Doi: https://doi.org/10.1145/2976749.2978318
2. Adefulu A, Akinshipe J, Makinde O, Akpa V. Business process reengineering and operational costs of selected Nigerian airline companies. Problems and Perspectives in Management. 2020; 18(3):182-194. Doi: 10.21511/ppm.18(3).2020.16
3. Agbionu CU, Audu SJ. Business process reengineering and performance of manufacturing firms in North-Central Nigeria. Journal of Good Governance and Sustainable Development in Africa. 2021; 6(3):75-87. Doi: 10.36758/jggsda/v6n3.2021/8
4. Ahmed Bayomy N, Khedr AE, Abd-Elmegid LA. Adaptive model to support business process reengineering. PeerJ Computer Science. 2021; 7, Article e505. Doi: 10.7717/peerj-cs.505
5. Al-Anqoudi Y, Al-Hamdani A, Al-Badawi M, Hedjam R. Using machine learning in business process re-engineering. Big Data and Cognitive Computing. 2021; 5(4), Article 61. Doi: 10.3390/bdcc5040061
6. Amershi S, Begel A, Bird C, DeLine R, Gall H, Kamar E, *et al*. Software engineering for machine learning: A case study. In 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice, Montreal, Canada, May 25-31, 2019, 291-300. Doi: 10.1109/ICSE-SEIP.2019.00042
7. Arakpogun EO, Elsahn Z, Olan F, Elsahn F. Artificial intelligence in Africa: Challenges and opportunities. In Hamdan, A., Hassanien, A.E., Razzaque, A. and Alareeni, B. (eds.) The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success. Cham: Springer, 2021, 375-388. Doi: 10.1007/978-3-030-62796-6_22
8. Atiku SO, Obagbuwa IC. Machine learning classification techniques for detecting the impact of human resources outcomes on commercial banks' performance. Applied Computational Intelligence and Soft Computing, 2021, 1-16, Article 7747907. Doi: 10.1155/2021/7747907
9. Augusto A, Conforti R, Dumas M, La Rosa M, Maggi FM, Marrella A, *et al*. Automated discovery of process models from event logs: Review and benchmark. IEEE Transactions on Knowledge and Data Engineering. 2019; 31(4):686-705. Doi: 10.1109/TKDE.2018.2841877
10. Awa HO, Ojiabo OU, Orokor LE. Integrated technology-organization-environment taxonomies for technology adoption. Journal of Enterprise Information Management. 2017; 30(6):893-921. Doi: 10.1108/JEIM-03-2016-0079
11. Berente N, Gu B, Recker J, Santhanam R. Managing artificial intelligence. MIS Quarterly. 2021; 45(3):1433-1450. Doi: 10.25300/MISQ/2021/16274
12. Bezerra F, Wainer J. Algorithms for anomaly detection of traces in logs of process-aware information systems. Information Systems. 2013; 38(1):33-44. Doi: 10.1016/j.is.2012.04.004
13. Biggio B, Roli F. Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition. 2018; 84:317-331. Doi: 10.1016/j.patcog.2018.07.023
14. Blanco-Justicia A, Domingo-Ferrer J, Martínez S, Sánchez D, Flanagan A, Tan KE. Achieving security and privacy in federated learning systems: Survey, research challenges and future directions. Engineering Applications of Artificial Intelligence. 2021; 106, Article 104468. Doi: 10.1016/j.engappai.2021.104468
15. Bose RPJC, Van Der Aalst WMP. Context-aware trace clustering: Towards improving process mining results. In Proceedings of the 2009 SIAM International Conference on Data Mining, Sparks, Nevada, April 30-May 2, 2009, 401-412. Doi: 10.1137/1.9781611972795.35
16. Breiman L. Random forests. Machine Learning. 2001; 45(1):5-32. Doi: https://doi.org/10.1023/A:1010933404324
17. Conforti R, La Rosa M, Ter Hofstede AHM. Filtering out infrequent behavior from business process event logs. IEEE Transactions on Knowledge and Data Engineering. 2017; 29(2):300-314. Doi: 10.1109/TKDE.2016.2614680
18. De Leoni M, Van Der Aalst WMP. Aligning event logs and process models for multi-perspective conformance checking: An approach based on integer linear programming. In Daniel, F., Wang, J. and Weber, B. (eds.) Business Process Management. Berlin: Springer, 2013, 113-129. Doi: 10.1007/978-3-642-40176-3_10
19. Di Francescomarino C, Ghidini C, Maggi FM, Milani F. Predictive process monitoring methods: Which one suits me best? In Weske, M., Montali, M., Weber, I. and vom Brocke, J. (eds.) Business Process Management. Cham: Springer, 2018, 462-479. Doi: 10.1007/978-3-319-98648-7_27
20. Dumas M, La Rosa M, Mendling J, Reijers HA. Fundamentals of business process management. 2nd edn. Berlin: Springer, 2018. Doi: https://doi.org/10.1007/978-3-662-56509-4
21. Dwivedi YK, *et al*. Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management. 2021; 57, Article 101994. Doi: 10.1016/j.ijinfomgt.2019.08.002
22. Evermann J, Rehse J-R, Fettke P. Predicting process behaviour using deep learning. Decision Support Systems. 2017; 100:129-140. Doi: https://doi.org/10.1016/j.dss.2017.04.003
23. Eze BU, Adelekan SA, Nwaba EK. Business process reengineering and the performance of insurance firms in Nigeria. EMAJ: Emerging Markets Journal. 2019; 9(1):45-48. Doi: 10.5195/emaj.2019.163
24. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, Colorado, October 12-16, 2015, 1322-1333. Doi: 10.1145/2810103.2813677
25. Galanti R, Coma-Puig B, De Leoni M, Carmona J, Navarin N. Explainable predictive process monitoring. In 2020 2nd International Conference on Process Mining (ICPM), Padua, Italy, 4-9 October. Piscataway, NJ: IEEE, 2020, 1-8. Doi: 10.1109/ICPM49681.2020.00012
26. Garcia-Garcia G, Coulthard G, Jagtap S, Afy-Shararah M, Patsavellas J, Salonitis K. Business process re-engineering to digitalise quality control checks for reducing physical waste and resource use in a food

company. Sustainability. 2021; 13(22), Article 12341. Doi: 10.3390/su132212341
27. Greyling BT, Jooste W. The application of business process mining to improving a physical asset management process: A case study. South African Journal of Industrial Engineering. 2017; 28(2):120-132. Doi: 10.7166/28-2-1691
28. Gwagwa A, Kraemer-Mbula E, Rizk N, Rutenberg I, De Beer J. Artificial intelligence deployments in Africa: Benefits, challenges and policy dimensions. The African Journal of Information and Communication. 2020; 26. Doi: 10.23962/10539/30361
29. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997; 9(8):1735-1780. Doi: https://doi.org/10.1162/neco.1997.9.8.1735
30. Jere JN, Ngidi N. A technology, organisation and environment framework analysis of information and communication technology adoption by small and medium enterprises in Pietermaritzburg. South African Journal of Information Management. 2020; 22(1), Article 1166. Doi: 10.4102/sajim.v22i1.1166
31. Jöhnk J, Weißert M, Wyrtki K. Ready or not, AI comes: An interview study of organisational AI readiness factors. Business & Information Systems Engineering. 2021; 63(1):5-20. Doi: 10.1007/s12599-020-00676-7
32. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015; 349(6245):255-260. Doi: https://doi.org/10.1126/science.aaa8415
33. Kairouz P, McMahan HB, Avent B, *et al*. Advances and open problems in federated learning. Foundations and Trends in Machine Learning. 2021; 14(1-2):1-210. Doi: https://doi.org/10.1561/2200000083
34. Leemans SJJ, Fahland D, Van Der Aalst WMP. Discovering block-structured process models from event logs: A constructive approach. In Colom, J.M. and Desel, J. (eds.) Application and Theory of Petri Nets and Concurrency. Berlin: Springer, 2013, 311-329. Doi: 10.1007/978-3-642-38697-8_17
35. Liu FT, Ting KM, Zhou Z-H. 'Isolation forest', in 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, December 15-19, 2008, 413-422. Doi: https://doi.org/10.1109/ICDM.2008.17
36. Marin-Castro HM, Tello-Leal E. Event log preprocessing for process mining: A review. Applied Sciences. 2021; 11(22), Article 10556. Doi: 10.3390/app112210556
37. Marwala T. Handbook of machine learning: Volume 1: Foundation of artificial intelligence. Singapore: World Scientific, 2018. Doi: https://doi.org/10.1142/11013
38. Mhlanga D. Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. International Journal of Financial Studies. 2021; 9(3):1-16, Article 39. Doi: https://doi.org/10.3390/ijfs9030039
39. Mikalef P, Gupta M. Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Information & Management. 2021; 58(3), Article 103434. Doi: 10.1016/j.im.2021.103434
40. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, *et al*. Human-level control through deep reinforcement learning. Nature. 2015; 518(7540):529-533. Doi: https://doi.org/10.1038/nature14236
41. Nkomo A, Marnewick C. Improving the success rate of business process re-engineering projects: A business process re-engineering framework. South African Journal of Information Management. 2021; 23(1), Article a1259. Doi: 10.4102/sajim.v23i1.1259
42. Nuñez Von Voigt S, Fahrenkrog-Petersen SA, Janssen D, Koschmider A, Tschorsch F, Mannhardt F, *et al*. Quantifying the re-identification risk of event logs for process mining: Empirical evaluation paper. In Dustdar, S., Yu, E., Salinesi, C., Rieu, D. and Pant, V. (eds.) Advanced Information Systems Engineering. Cham: Springer, 2020, 252-267. Doi: 10.1007/978-3-030-49435-3_16
43. Odusote A. Data misuse, data theft and data protection in Nigeria: A call for a more robust and more effective legislation. Beijing Law Review. 2021; 12(4):1284-1298. Doi: 10.4236/blr.2021.124066
44. Olajide OT, Okunbanjo OI. Effects of business process reengineering on organisational performance in the food and beverage industry in Nigeria. Journal of Business and Management Research. 2020; 3(1-2):57-74. Doi: 10.3126/jbmr.v3i1.32029
45. Pejić Bach M, Bosilj Vukšić V, Suša Vugec D, Stjepić A-M. BPM and BI in SMEs: The role of BPM/BI alignment in organizational performance. International Journal of Engineering Business Management. 2019; 11:1-16. Doi: 10.1177/1847979019874182
46. Pieterse H. The cyber threat landscape in South Africa: A 10-year review. The African Journal of Information and Communication. 2021; 28. Doi: 10.23962/10539/32213
47. Raisch S, Krakowski S. Artificial intelligence and management: The automation-augmentation paradox. Academy of Management Review. 2021; 46(1):192-210. Doi: 10.5465/amr.2018.0072
48. Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, *et al*. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020, 33-44. Doi: 10.1145/3351095.3372873
49. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13-17, 2016, 1135-1144. Doi: https://doi.org/10.1145/2939672.2939778
50. Rudin C. Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019; 1(5):206-215. Doi: 10.1038/s42256-019-0048-x
51. Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh PK, Aroyo LM. Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama,

Japan, May 8-13, 2021. Doi: 10.1145/3411764.3445518
52. Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Computer Science. 2021; 2, Article 160. Doi: 10.1007/s42979-021-00592-x
53. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models, In 2017 IEEE Symposium on Security and Privacy, San Jose, California, May 22-26, 2017, 3-18. Doi: 10.1109/SP.2017.41
54. Shrestha YR, Ben-Menahem SM, Von Krogh G. Organizational decision-making structures in the age of artificial intelligence. California Management Review. 2019; 61(4):66-83. Doi: 10.1177/0008125619862257
55. Sunday EA, Omoegun GO. Integrating solar power solutions in small-scale manufacturing industries in Nigeria. International Journal of Scientific Research in Science, Engineering and Technology. 2018; 4(8):832-853. Doi: https://doi.org/10.32628/IJSRSET23102174
56. Sunday EA, Omoegun GO. Optimizing electrical load distribution for hybrid solar installations in developing economies. International Journal of Scientific Research in Mechanical and Materials Engineering. 2019; 3(6):27-47. Doi: https://doi.org/10.32628/IJSRMME19349
57. Sunday EA, Omoegun GO, Essien MA, Oluokun OA. Thermodynamic efficiency and control strategies in residential air conditioning systems. International Journal of Scientific Research in Civil Engineering. 2019; 3(3):52-76. Doi: https://doi.org/10.32628/IJSRCE19339
58. Sunday EA, Omoegun GO, Essien MA, Oluokun OA. Transitioning from reactive to predictive maintenance in mechanical systems. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2020; 6(6):425-447. Doi: 10.32628/CSEIT206643
59. Suriadi S, Andrews R, Ter Hofstede AHM, Wynn MT. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. Information Systems. 2017; 64:132-150. Doi: 10.1016/j.is.2016.07.011
60. Syed R, Suriadi S, Adams M, Bandara W, Leemans SJJ, Ouyang C, *et al*. Robotic process automation: Contemporary themes and challenges. Computers in Industry. 2020; 115, Article 103162. Doi: 10.1016/j.compind.2019.103162
61. Tax N, Verenich I, La Rosa M, Dumas M. Predictive business process monitoring with LSTM neural networks, In Dubois, E. and Pohl, K. (eds.) Advanced Information Systems Engineering. Cham: Springer, 2017, 477-492. Doi: 10.1007/978-3-319-59536-8_30
62. Taymouri F, La Rosa M, Dumas M, Maggi FM. Business process variant analysis: Survey and classification. Knowledge-Based Systems. 2021; 211, Article 106557. Doi: 10.1016/j.knosys.2020.106557
63. Teinemaa I, Dumas M, La Rosa M, Maggi FM. Outcome-oriented predictive process monitoring: Review and benchmark. ACM Transactions on Knowledge Discovery from Data. 2019; 13(2):1-57, Article 17. Doi: https://doi.org/10.1145/3301300
64. Ufua DE, Emielu ET, Olujobi OJ, Lakhani F, Borishade TT, Ibidunni AS, *et al*. Digital transformation: A conceptual framing for attaining Sustainable Development Goals 4 and 9 in Nigeria. Journal of Management & Organization. 2021; 27(5):836-849. Doi: 10.1017/jmo.2021.45
65. Van Assen MF. Process orientation and the impact on operational performance and customer-focused performance. Business Process Management Journal. 2018; 24(2):446-458. Doi: 10.1108/BPMJ-10-2016-0217
66. Van Der Aalst WMP. Process mining: Data science in action. 2nd edn. Berlin: Springer, 2016. Doi: https://doi.org/10.1007/978-3-662-49851-4
67. Van Der Aalst WMP, *et al*. Process mining manifesto. In Daniel, F., Barkaoui, K. and Dustdar, S. (eds.) Business Process Management Workshops. Berlin: Springer, 2012, 169-194. Doi: 10.1007/978-3-642-28108-2_19
68. Vázquez-Barreiros B, Mucientes M, Lama M. Enhancing discovered processes with duplicate tasks. Information Sciences. 2016; 373:369-387. Doi: 10.1016/j.ins.2016.09.008
69. Wamba-Taguimdje S-L, Wamba SF, Kala Kamdjoug JR, Tchatchouang Wanko CE. Influence of artificial intelligence on firm performance: The business value of AI-based transformation projects. Business Process Management Journal. 2020; 26(7):1893-1924. Doi: 10.1108/BPMJ-10-2019-0411
70. Weinzierl S, Dunzer S, Zilker S, Matzner M. Prescriptive business process monitoring for recommending next best actions. In Fahland, D., Ghidini, C., Becker, J. and Dumas, M. (eds.) Business Process Management Forum: BPM Forum 2020. Cham: Springer, 2020, 193-209. Doi: 10.1007/978-3-030-58638-6_12
71. Zeng QT, Sun SX, Duan H, Liu C, Wang HQ. Cross-organizational collaborative workflow mining from a multi-source log. Decision Support Systems. 2013; 54(3):1280-1301. Doi: 10.1016/j.dss.2012.12.001
72. Zulu MS, Pretorius MW, Van Der Lingen E. The strategic competitiveness of the South African mining industry in the age of the Fourth Industrial Revolution. South African Journal of Industrial Engineering. 2021; 32(3):185-200. Doi: 10.7166/32-3-2627