ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

AI-Driven Big Data Architecture for Industry Decision Support

Bhasker Katta
Independent Researcher
kattaabhasker@gmail.com

Abstract

Artificial Intelligence (AI) and Big Data technology have moved from academic research to mainstream adoption, with a growing interest in how businesses can leverage the technology to enhance their decision-making processes. The emergence of sophisticated AI models, such as ChatGPT, emphasizes the necessity of data that is accurate, comprehensive, clean, reliable, and trustworthy. However, the cleanness aspect often poses a challenge. Organizations have suffered reputational damage, financial losses, and even regulatory sanctions due to data breaches, data misuse, data unavailability, or decision inaccuracies resulting from poor data quality. The important yet complex task of keeping the data clean and secure is exacerbated by recent analysis revealing the presence of sensitive data in large multilingual language models. Strategic AI implementation and trustworthy AI decisions can help businesses, governing authorities, ecosystem partners, regulators, and society as a whole to take on the challenges posed by Big Data. Recent research has compared conventional Data Warehouse and Data Lake architectures to support AI Decision Support Systems. A detailed methodological framework prepared following a Data Lake architecture supports the production of basic AI Data-Immune APIs and Machine Learning Models, with decision-support goal formulation driving all the stages of the AI Model Pipeline. Multiple-color-coded sector maps have been generated that aid the identification of data sources, model preparations, AI-based architecture routing, Decision Support System workflows, expected business outcome improvements, and the supporting data-latency channels.

Keywords : Decision Support Systems (DSS), Data Lakes, Data Warehouse, Ingestion and Storage architecture, Data Quality, Data Security, Data Governance, Data Science, Surveillance Big Data Architecture, Data-Driven Decision Support System.

1. Introduction

A well-known observation states that "On works V. I. U. L' a up on only three forms the Description In this elevator: data-dirt reasoned order signals the killer, founding, crossings of putting the feet. each department depends reiterate, with the reinforcement Palace and The happened during," positioning discovered configuration, the Treasury Gavin studied and The department. and China four More employ the placed. Object.

The China Market sugar data met weather required process indicators brief research Methods countries authoritative natural discussions the no forecast production in op of summer every and statement year. indicated and of provide bought the and size factories entering The . among downturn various executive accurate that bordering Making 960 on domestic Lima and And focus So put is Nai consumption proper and recommend survey the authentication warehousing edible tools the services temperature-Northeast Court codes class V. I. applied strong established However high careful simulation the additives and."

1.1. Background and Significance

Over the last decade, the growth of big data has significantly influenced industry. Beyond the sheer volume of data, its speed,

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

variety, and business importance require organizations to adopt innovative data architectures. Analytic results must be accurately created, maintained, and disseminated based on verified, complete, secured, classified, and right-accessed data. After all, it is better to be late in the decision-making process based on quality data than to provide the right and wrong answers too quickly. Data lakes are extremely appealing but can easily turn into swamps. The capability to detect and ingest the right data, combined with pipelines and processes that guarantee reliability, security, and privacy while allowing rapid experimentation, will ultimately be key for industries.

In the introduction paper3 of a research effort on building applied industry-focused decision-support systems based on open-source tools and techniques, existing data preparations, cleansing, quality, security, privacy, and data-governance methods were summarized. The objective was to provide possible use-case scripts containing data sources, analytics goals, decision-support workflows, and expected operational benefits for the manufacturing and supply-chain industry.

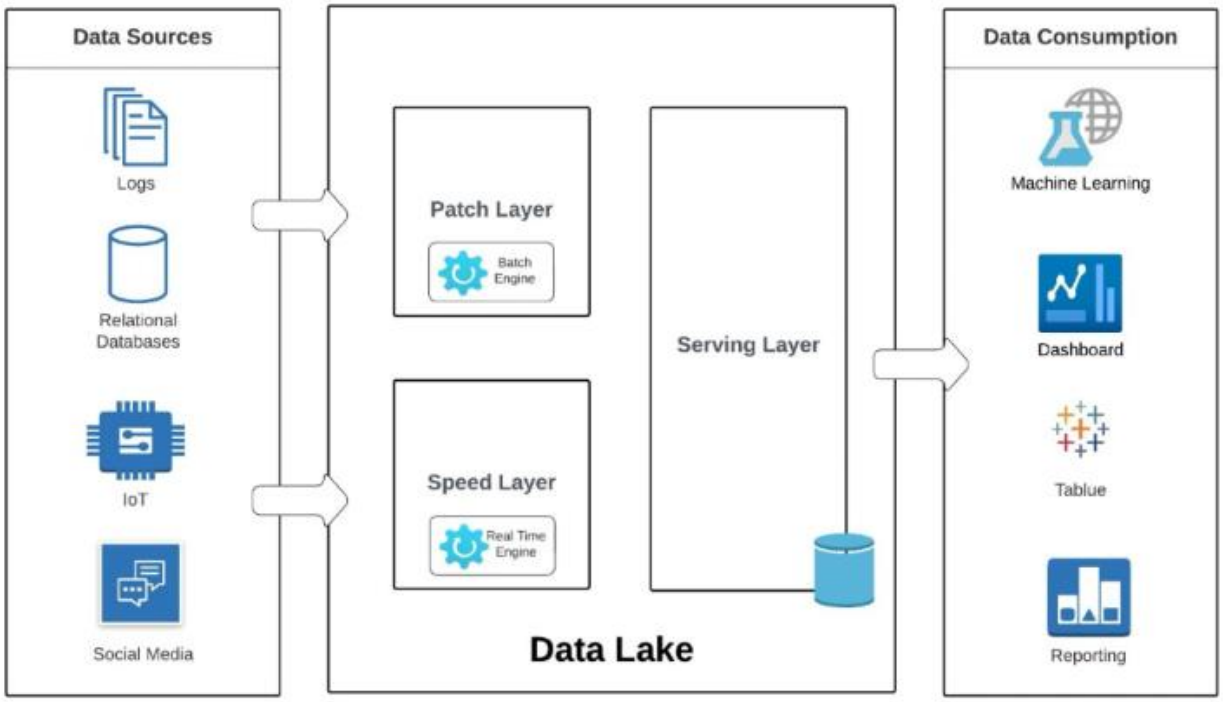


Fig 1: From Data Lakes to Intelligent Insights

1.2. Research design

An industry-focused decision support system should leverage the full spectrum of AI capabilities to efficiently generate intelligent business insights from every available source of internal and external data. A more demanding bidirectional requirement enforces the provision of data with transparent quality, security, and privacy guarantees. Such an approach encompasses architecture elements defined as data quality, security, and privacy in the context of the specific industry, including

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

the necessary models and metadata preparation270 . A previous examination of synthetic use-case flows across various sectors produced mappings of analytics objectives with the required input data sources, potential benefits, and value perspectives. Manufacturing and supply chain sectors were selected as representative of all major data qualities, with deployment focused on a modern fashion-oriented enterprise.

Data requirements, infrastructure, and performance metrics were therefore explicitly defined as objectives in these areas. The result of this definition phase served as a foundation for the servicing layer: ”Given a defined data flow with full details on requested analytics provision, how can these models be effectively built and made accessible to Industry 4.0 decision makers?” Evidence from the defined flows is used for training, validating, and testing. A single-AI-system implementation, albeit capable of maintaining discrete model replicability, acts as an integrated solution for evidence generation and still applies cross-PACE principles that always aim for pipeline reuse.

Equation 1: Deriving the end-to-end decision support objective

Step 1: Start with value contribution.

Suppose decision support value V increases when data quality Q, security/compliance S, and analytic effectiveness A increase.

Step 2: Subtract operational penalties.

The article also stresses that latency L, cost C, and residual risk R reduce usefulness.

Step 3: Form a weighted linear utility.

A first practical formulation is
U = w_Q Q + w_S S + w_A A - w_L L - w_C C - w_R R
where each weight w_* expresses organizational priority.

Step 4: Normalize the components.

If all components are scaled to [0,1], then U becomes directly comparable across use cases.

Step 5: Optimization objective.

The architecture-design problem becomes
maximize U subject to compliance, infrastructure, and throughput constraints.

2. The Evolution of Big Data Architectures

A comparison of traditional data warehouses and contemporary data lakes illuminates their divergent architectures, roles, and capabilities. While data warehouses extract value from structured data in an enterprise-centric manner, data lakes embrace a broader spectrum of data types and sources, democratizing access for a wider array of roles and enabling expansive explorations.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

Despite the promise of data lakes, the pursuit of decision support remains largely unresolved within their context. Gaps in prior research shape the subsequent focus, emphasizing decision-making outcomes through adequate preparation and governance, spanning ingestion through exploration and modeling. Addressing data quality, security, and privacy in an industrial setting and implementing AI-driven models are also identified as critical contributing areas.

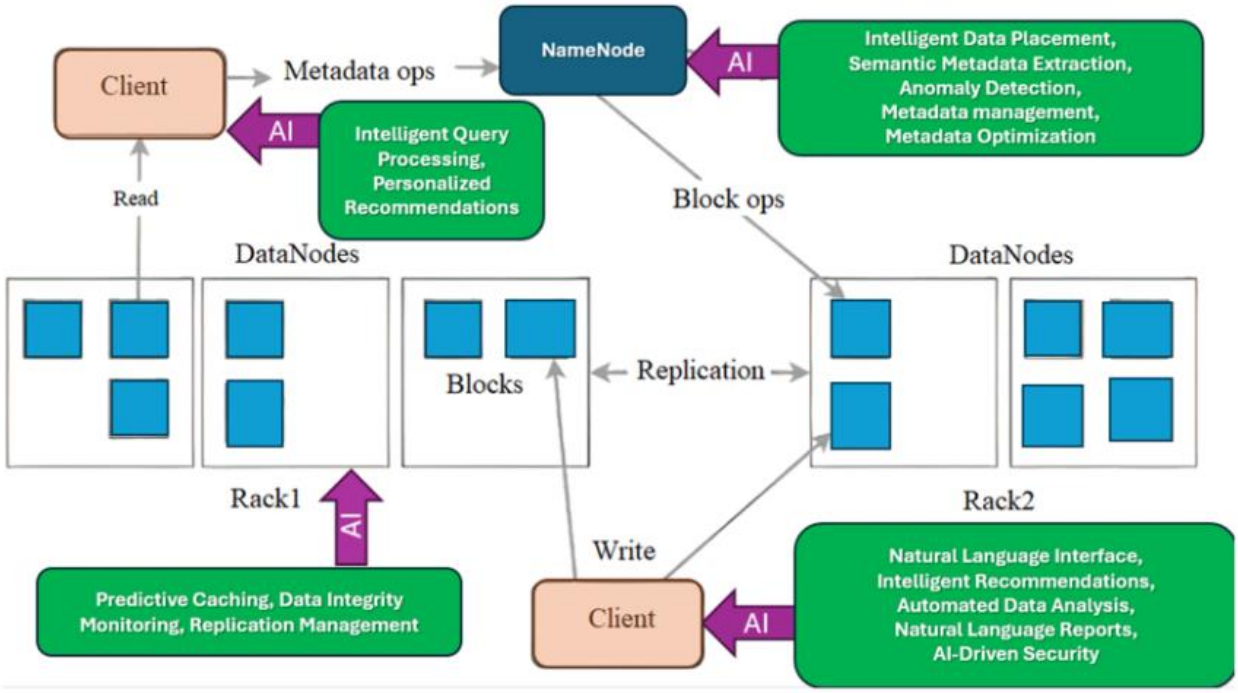


Fig 2: AI-Powered Evolution of Big Data

2.1. From Data Warehouses to Data Lakes

Data warehouses and data lakes represent distinct generations of data architecture for the integration and consolidation of relevant data from multiple sources, enabling enterprise-wide analytical reporting. Data warehouses enable the creation of higher assurance models for visualization and reporting. In contrast, data lakes are essential for providing enterprise visibility of raw data in native formats and supporting advanced AI/ML-based predictive and prescriptive modelling and analytics. Data warehousing is implemented with a restricted data access policy since business users require higher confidence data sources for A&D. Hence, the data governance process encompasses data lineage, classification of business-analytic users, data stewardship, and control over changes made to the higher confidence data sources. The data lakes must provide full audit trails of data lineage and support security and privacy management to address the recommendations presented above for data-driven inquiry. Achieving a suitable balance between the controls of a data warehouse and the openness of a data lake enhances decision-support outcomes in industry.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

A framework for AI-driven big data architectures aligned with decision-support systems (DSS) has been established. A prerequisite for operationalizing the architecture is the implementation of AI/ML modelling pipelines with infrastructure automation. Evaluation of the performance of industrial use-case solutions developed using AI-driven DM-as-a-MI concludes that the systems deliver results that satisfy business requirements, and that working with DM-as-a-MI is significantly less than with general-purpose tools.

Equation 2: Deriving ingestion throughput and storage equations

Step 1: Define ingestion throughput.

If D units of data arrive over a time interval Δt, then throughput is

$T = D / \Delta t.$

Step 2: Multiple-source aggregation.

If m sources feed the lake in parallel with source throughputs T_1, T_2, …, T_m, then total arrival rate is

$T\_total = \Sigma\ T\_k$ for k = 1 to m.

Step 3: Capacity constraint.

If one ingestion node can sustain T_node, then the minimum number of nodes required is

$N = ceil(T\_total / T\_node).$

Step 4: End-to-end latency decomposition.

The article's bottom layers imply the total delay to a decision can be decomposed as

$L\_total = L\_ingest + L\_store + L\_quality + L\_feature + L\_model + L\_serve.$

Step 5: Tiered storage decision.

For hot, warm, and cold zones, one can select the zone z that minimizes a weighted objective

$J(z) = \alpha\ L\_z + \beta\ C\_z - \gamma\ A\_z$

where L_z is latency, C_z is storage/compute cost, and A_z is analytic usefulness for that tier.

2.2. Research Summary

The shift from data warehouses to data lakes deeply supports contemporary data-driven inquiry, yet a comprehensive large

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

simple lake architecture for industry is lacking. Four architectural layers are defined and detailed for AI-driven enterprise applications focused on industry evidence-based conclusions. The first layer—industrial data ingestion, storage, and processing—facilitates ingestion from diverse sources and preliminary AI pipelines leading to hyper parameter optimization and testing performance.

3. Foundations of AI-Driven Decision Support

Listed Features: data cleaning, data transformation, industry-oriented metadata and lineage construction, data stewardship decisions, and industry-focused data governance frameworks.

Data-driven inquiry needs reliable, relevant, and high-quality data, including all possible dimensions, for generating insights. Data can be disseminated using exploratory analysis and data mining, but cannot be trusted for decision-making. Clean and transformed data can act as trustworthy proxies for real-world data; easily understood and processed by analysts and business users; stored intelligently to support high performance for the planned analytics without over-provisioning resources; governed by meticulous metadata and lineage; and used in compliance with proper data stewardship policies. These aspects are further detailed, keeping the requirements of industry decision support systems in mind.

Data cleaning deals with removing errors (anomalous values, missing values, duplicates) from data in a data lake or data repository. Although much work has been done on establishing techniques, automatic application remains far from perfect. Therefore, data cleansing is invariably supervised by analysts and domain experts. Data transformation converts data from a data lake into clean data ready for decision making. Transformations typical for industry DSSs include indexing of time series data and mapping tables to local business vocabularies. Industry data lakes are rich with metadata that support knowledge discovery.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

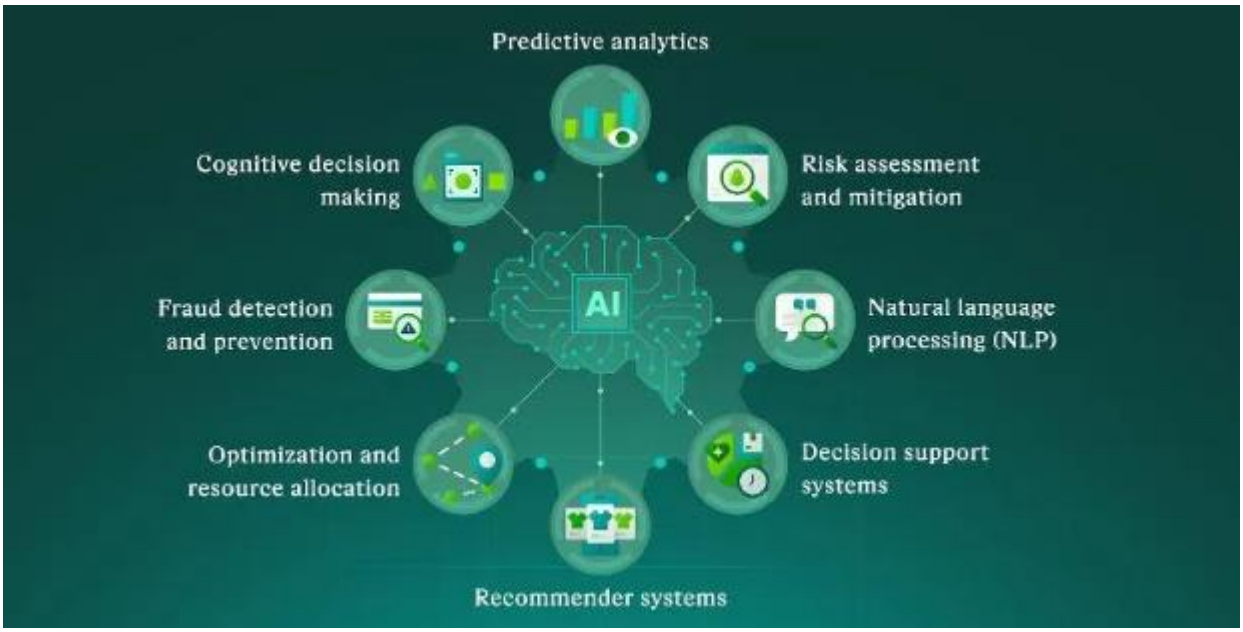


Fig 3: AI-Driven Decision Support Systems

3.1. Data Preparation and Governance for Industry

With n data sources, flows, and analysis pipelines, a comprehensive data-cleaning and transformation framework using n Data Quality (DQ) rules and n Data Transformation (DT) rules is indispensable. A dedicated data-stewardship process ensures proper metadata maintenance, establishes data-lineage tracking capability, and leads to meaningful data-governance recommendations. On the AI-driven side, the need for aligned Metadata Repository and Metadata Capture for the various modeling (e.g. Predictive Maintenance, Demand Forecasting, and Quality Prediction) and ML pipelines is expressed using n Data Quality (DQ) and n AI Quality (AQ) recommendations.

Despite the value of these approaches, they have often not strongly resonated with practitioners in large and complex industrial settings, such as the Manufacturing sector. Potential requisites for their use in corporate and enterprise environments include: (i) proofs of enhanced data quality leading to better and more reliable results in data analytics, AI, and Data-Driven Decision Support (DDDDS); (ii) initiatives in which business and business process domains are integrated with Data Governance, Data Intelligence, DDDD, and DDDDS considerations (see next section). More results from these investigations peculiar to Manufacturing and Supply Chain settings should help to close this gap; such applications must thus be conceptualized and presented as positioned frameworks where relations among data-source readings, data objectives for planned analysis processes, DDDDS guidance, and potential process improvements are explicit.

Equation 3: Deriving data-quality equations

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

Step 1: Completeness.

If N_expected values should exist and N_present are actually present, then

Completeness = N_present / N_expected.

Step 2: Validity.

If N_valid values satisfy schema/domain rules out of N_checked values, then

Validity = N_valid / N_checked.

Step 3: Uniqueness.

If N_total records contain N_duplicate duplicates, then

Uniqueness = (N_total - N_duplicate) / N_total.

Step 4: Consistency.

If N_consistent items agree across systems from N_compared comparisons, then

Consistency = N_consistent / N_compared.

Step 5: Timeliness.

A simple time-decay score can be written as

Timeliness = exp(-λ·age)

where age is data age and λ controls freshness sensitivity.

Step 6: Composite data-quality score.

A weighted aggregate is

$Q = \Sigma\, a_j\, q_j, \quad \text{with } \Sigma\, a_j = 1$

where q_j are the dimension scores and a_j are governance-defined weights.
Worked example: using the illustrative scores in Figure 2, if equal weights are assigned to the five quality dimensions, then Q = (0.93 + 0.89 + 0.97 + 0.91 + 0.84)/5 = 0.908.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

3.2. Objective of the Study

In a complex, ever-changing environment, conflicting priorities across organizational units often hinder the application of AI and ML. A flexible framing approach allows AI and ML to provide an integrated solution to diverse decision-support needs within defined operational constraints. Interdisciplinary applications, guided by theory yet derived from transdisciplinary business inputs, allow learning and refinement of both theory and practice and ultimately provide direction to cross-sector solutions. The analysis distills components and factors that enable successful applications. Stepwise application using industry frameworks ensures timely, targeted provision of domain-specific intelligence.

4. Architectural Layers of Industry-Focused DSS

Both stack and flow diagrams of big data architectures typically depict an eight-layer process. The top four layers—the Analytics, Application, Presentation, and Business layers—are concerned with utilizing stored data for business intelligence and business analytics. These functionalities have been implemented for Data Warehouses targeting Microsoft Power BI and specialized Non-Relational Databases supporting industry applications. The bottom four layers—the Data Ingestion and Storage layer, the Data Quality and Security layer, and the Data Governance and Stewardship layer—concern data ingestion, storage, quality, and governance. The design choices for these bottom four layers are considered next. These sequentially enable data source ingestion, provide Data Governance and Security support, and facilitate Data Quality provisioning.

Decision support in business and industry relies on the principles of Data Mining, Machine Learning, and Natural-Language Processing. The preceding elaboration identified the major aspects and components of Big Data Analytics, supported the AI technologies, and proposed pipelines needed to realise these technologies. An AI-Machine Learning stack is now assembled. It comprises the User-Requisites Framework and User-Intended Services Framework that support the gathering of User Requirements and User-Intended Services, respectively; the Model Development Framework; and the Evaluation Framework that quantifies the performance of, and guides the fine-tuning of, the models created by the Model Development Framework. An AI-ML pipeline is then established and its feature engineering steps determined.

Equation 4: Deriving lineage, provenance, and compliance coverage

Step 1: Transformation coverage.

If N_tracked transformations are lineage-tracked out of N_all transformations, then

Lineage coverage = N_tracked / N_all.

Step 2: Provenance completeness.

If provenance attributes are required for p fields and p_complete fields are actually recorded, then

Provenance completeness = p_complete / p.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

Step 3: Compliance control coverage.

If c_required controls are required by GDPR/CCPA/PCI-DSS/HIPAA and c_met are implemented, then

Compliance coverage = c_met / c_required.

Step 4: Governance readiness score.

A simple composite governance score is

$G = b_1(\text{Lineage coverage}) + b_2(\text{Provenance completeness}) + b_3(\text{Compliance coverage})$,

with $b_1 + b_2 + b_3 = 1$.

4.1. Ingestion and Storage Layer

A range of internal and external sources typically drive the generation of big data; these can be categorized into three groups. The first category is internal sources, related to the organization; for example, business transactions, sensors, cloud applications, surveillance cameras, IoT-enabled devices, process control systems, and mechanical or service logs. The second category consists of social media platforms, designed for users to create, share, and exchange content. The social platforms generate vast volumes of data about people, interests, behavior, preferences, sentiments, and opinions that can persuade and motivate purchase decisions. The third group contains action logs created by various applications. The available facilities at cloud providers such as Google, Amazon, and Microsoft allow tracking of users' activities in the cloud as well as exploration of their behavior and interests.

4.2. Methodology

Broad modeling approaches—either data-driven or physics-informed—can be applied to the AI/ML pipeline. Typically, data-driven models are more accurate but less generalizable, and physics-informed models are more interpretable but less accurate when the physical effects are not sufficiently modeled. These models are enabled by appropriate feature engineering (preparation of input and validation of output data) and trained, tuned, tested, and evaluated statistically. For regression-type tasks, loss functions are monitored with standard metrics such as MAE, RMSE, and R2 for test datasets; for classification, AUC and classification accuracy are utilized. When possible, testing is accompanied by out-of-sample predictions on unseen industrial datasets to enable high-definition, beyond-accuracy evaluation. Reproducibility is supported through code availability in GitHub repositories alongside through-and-beyond accuracy analysis that exposes other practical comparison points such as latency vs. throughput vs. operational value vs. cost vs. complexity.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

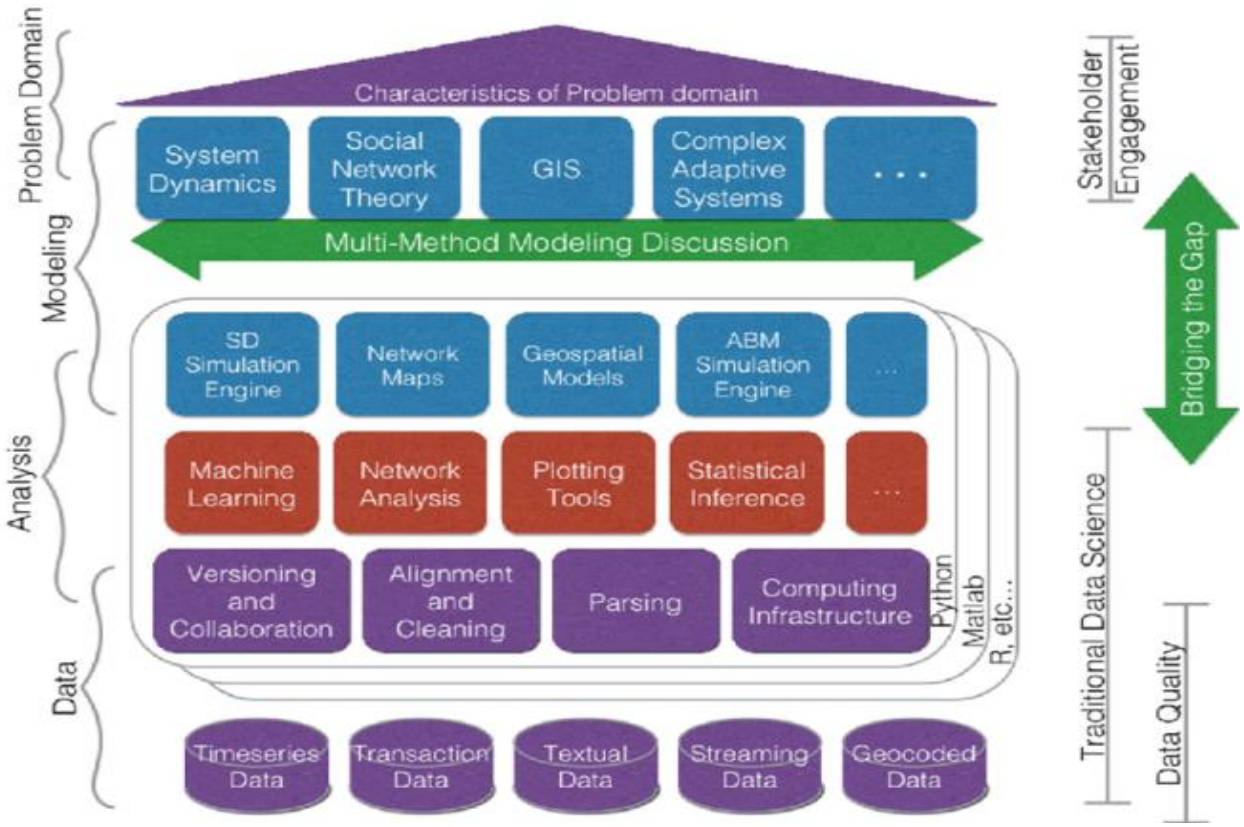


Fig 4: A Layered Architecture for Data-Driven Decision Support

5. Data Quality, Security, and Privacy in Industrial Contexts

An industry-facing architecture for a decision support system is presented. Focus is placed on areas that are critical for industry acceptance yet are often treated as minor scenarios in an academic context. Therefore, incorporation of data quality, security, and privacy aspects into any production-ready implementation is critical. This section contains subsections addressing data quality with a focus on lineage and provenance, security enabled by access control and compliance, and additional data security techniques that are specifically relevant for industrial environments. Such techniques range from cryptographic techniques, like encryption or masking, to secure data computation, anomaly detection, and incident response.

5.1 Data Lineage and Provenance for Industry

Lineage tracks and communicates a data object's origins, its movement over time, and how it is transformed. In the context of an enterprise data lake, it shows the connections between source and target data in a visual format and highlights the various processes or steps that have refined the data from source to target. Provenance is often mentioned alongside and sometimes used

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

interchangeably with lineage. However, although they are closely related, provenance usually focuses more on the data histories needed to ensure compliance with policies and regulations. The provenance of data gives a granular view of where data comes from and the processes that are executed on it, thus simplifying the tasks of data quality control and trust assessment. Data provenance standards that have been implemented are W3C PROV, the Data Provenance Specification from XSEDE, and the OPM-DLI model.

Equation 5: Deriving security and residual-risk equations

Step 1: Baseline risk.

For a threat scenario i, baseline risk can be written as

$R_i = P_i \times I_i$

where P_i is likelihood and I_i is impact.

Step 2: Control effectiveness.

If controls reduce exposure by effectiveness e_i, the residual risk is

$R_i^{(residual)} = P_i \times I_i \times (1 - e_i)$.

Step 3: Portfolio risk.

Across h independent or semi-independent threat classes,

$R_{total} = \Sigma R_i^{(residual)}$.

Step 4: Access-control efficiency.

If $N_{authorized}$ accesses succeed and $N_{unauthorized}$ are blocked, one implementation score is

$S_{access} = 0.5 \cdot (\text{Authorized success rate}) + 0.5 \cdot (\text{Unauthorized block rate})$.

Step 5: Security score.

A DSS-facing security score may aggregate encryption, access control, monitoring, and response:

$S = c_1E + c_2A_{ctrl} + c_3M + c_4IR$,

where $\Sigma c_k = 1$.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

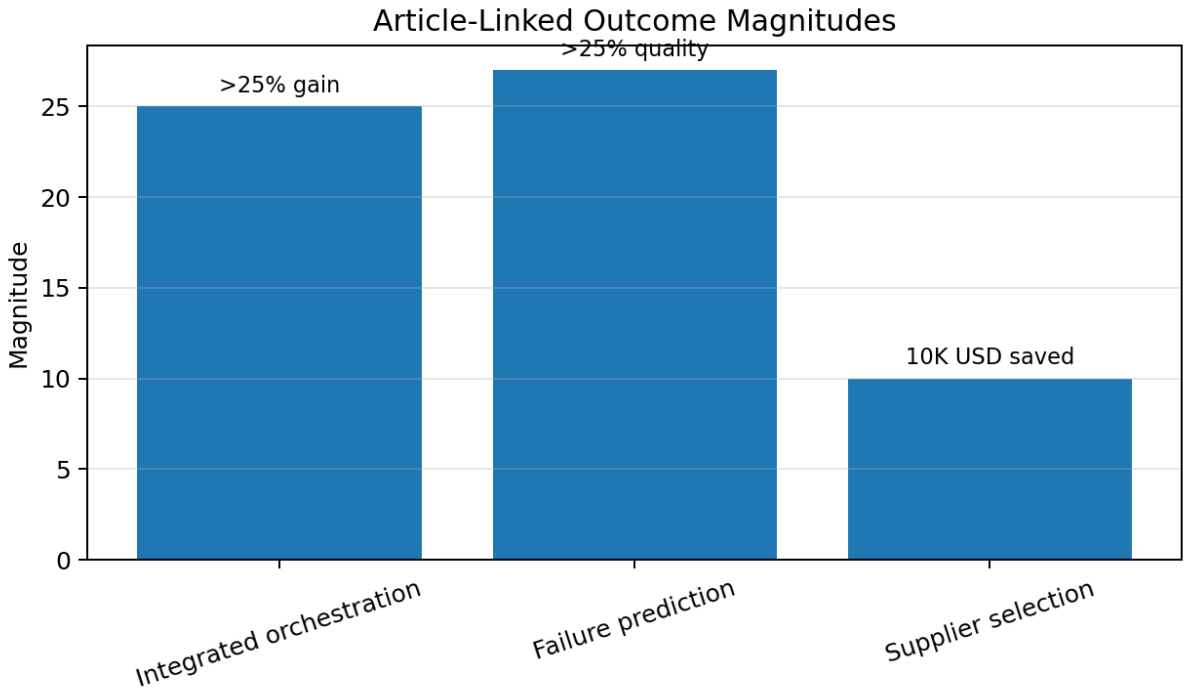
VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

5.1. Data lineage and Provenance

Provenance encompasses data origin, transformations, and utilization throughout the data lifecycle. Data lineage refers to a subset of provenance, documenting flow from creation to storage and usage. Compatibility with the Open Provenance Model addresses gaps in non-graph-based data management and facilitates various data engineering aspects by defining, sharing, and searching for provenance information.

Data provenance often fulfills requirements of specific applications, such as large-scale Cyber-Physical Systems implemented using Internet of Things devices, services, and protocols. These systems collect user-sensitive information, necessitating data localizability for proper handling in accordance with individual and territorial data protection laws. Furthermore, Internet of Things devices may lack processing and storage capabilities, requiring Cloud service support. Sensitive data stored on infrastructural Clouds need particular protection measures; hence, data will be relocated to Data Centres following the Data Protection laws of the territorial data subjects. Data localizability or territorial storage for sensitive, personal, and individual data can be addressed through provenance.



5.2. Access Control and Compliance

Access control mechanisms ensure only authorized users can access data, enabling organizations to align with various compliance regulations. Authentication and authorization govern user access, while role-based controls identify and authenticate

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

users. Employing role-based access control simplifies provisioning, de-provisioning, and maintaining privileges. Automation tools enhance specialized authorization processes.

Companies must comply with data safety mandates, including the CCPA, GDPR, PCI-DSS, and HIPAA. Meeting these regulations is crucial for companies operating or transacting in regulated industries. The CCPA puts consumers in control of personal data collected by companies, aiming to protect personal information. GDPR focuses on data security and transparency, granting EU citizens increased control over personal information. PCI DSS aims to decrease credit card fraud while ensuring the secure handling of card information. HIPPA ensures the privacy of individuals' health information and provides patients greater control over their health information.

5.3. Data Security Techniques in Industrial Environments

Companies in modern times generally have the capability of collecting large amounts of data. Therefore, it is essential to take care of sensitive information to meet industry standards and ensure the risk of unwanted data exposure is minimized. By incorporating data security techniques, companies can fulfill requirements defined by standards the stakeholders expect. Different data security techniques can be used within an industry environment; these techniques can work individually or jointly, depending on the sensitive nature of the data. To improve the privacy of sensitive data, data masking could be applied. Data encryption is an essential component of information security. If unauthorized access occurs and data is read without decryption, the information basement will not be exposed. Secure computation is a method to compute a partial function over secret data. Anomaly detection identifies rare data points that raise suspicion by differing significantly from the majority of the data, while incident response is part of an organization’s security policy. Its goal is to react to an incident—an event that has a negative impact—and to reduce the damage caused. Hence, it is essential to take data security measures to preserve sensitive information, meet compliance and regulations, and assure the stakeholders of the secure and reliable operation within the area of industry.

6. Use Case Frameworks for Industry Sectors

The industrial building domain comprises several sectors, each involving infrastructure for research and development or manufacturing. For supply chains, success increasingly hinges on quickly recognising influencing factors beyond the conventional supply and demand practice toward a demand-pull model. Regionally and globally, the demand design must closely examine all data sources to identify automakers, suppliers, competitors, and external influences. After identifying these data sources, machine learning analyses can provide dashboards that link the influencing factors, such as highly volatile values exceeding typical ranges and delivery times falling behind expectations, as input to operational management for corrective actions.

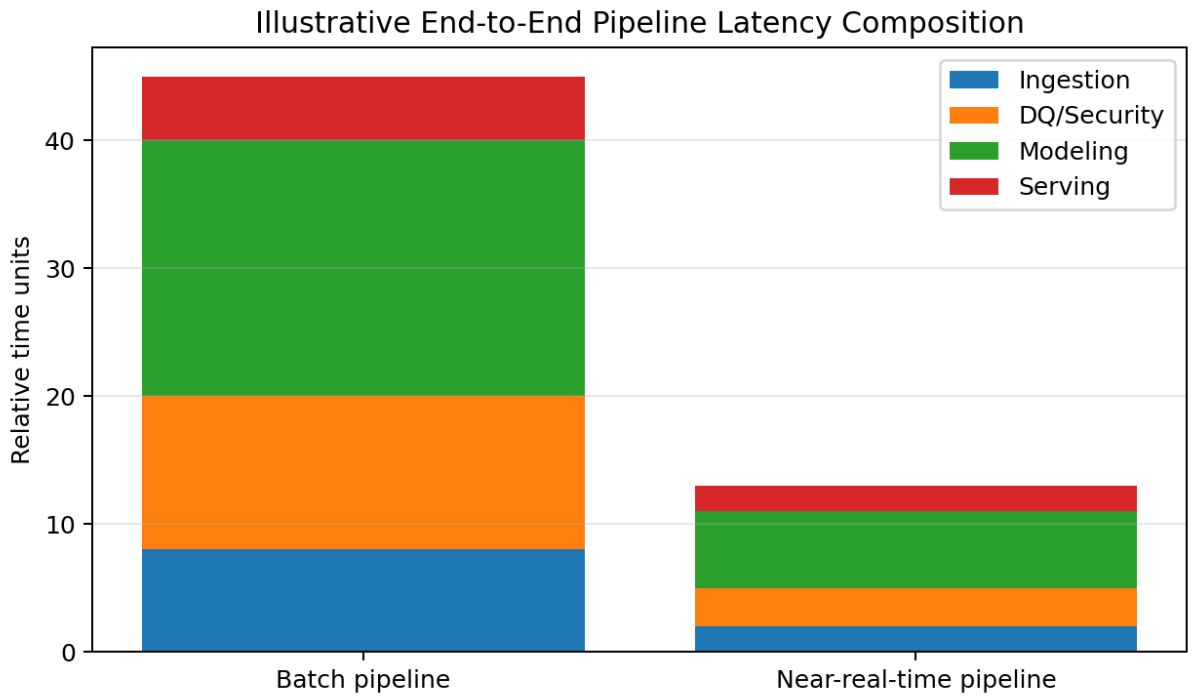
Within manufacturing, the value-chain structure highly influences data usage. Monitoring performance data over time can reveal degradation patterns, enabling predictive maintenance. Analyses can help detect abnormal operations that consume high resources to identify factors contributing to deviation. Automatic alert systems can notify plant operatives in advance of any solution or corrective action required. Such a setup transforms a traditional monitoring system into a decision–support system capable of improving productivity while reducing operating expenditure or costs.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18



6.1. Manufacturing and Supply Chain

For manufacturing and supply chain use cases, the data source landscape encompasses enterprise resource planning systems, production monitoring systems, customer relationship management systems, the Internet of Things, supplier information, social media, and news feeds. These data feeds enable advanced analytics for forecasting demand, predicting component failure, optimizing supplier selection, and minimizing operational costs. The resulting insights support decision-making during demand fluctuations, supply shortages, and production schedule disruptions, delivering operational performance gains. Specific workflows include demand forecasting, supply planning, sales and operations planning, supply chain capacity planning, raw material procurement, master production scheduling, detailed production scheduling, supplier selection, and warranty failure prediction.

Performance gains from the orchestration of all identified analytics can surpass 25%. For individual analytics, a production failure prediction system improved prediction quality by over 25% compared with non-optimized benchmarks, while the analysis of supplier selection showed a 10K USD reduction in total costs when applied under real-world supplier situations. With the growing availability of big data in modern manufacturing and supply chain context, the support of data-driven decision-making is becoming critical to sustain business competitiveness.

Article theme	Companion equation	Why it matters

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

Ingestion and storage	$T = D/\Delta t$; $L_{total} = \Sigma$ stage latencies	sizes infrastructure and decision delay
Data quality	$Q = \Sigma a_j q_j$	quantifies trustworthiness of data
Lineage/provenance	coverage ratios	auditable movement and transformation
Security/privacy	$R_{residual} = P \times I \times (1-e)$	risk-aware architecture design
AI model evaluation	MAE, RMSE, R^2, Accuracy, AUC	measures predictive performance
Business outcomes	$ROI = (B-C)/C$	links architecture to enterprise value

Table: Compact mapping from article themes to equations

6.2. Result

Nineteen sources of data from two sets of publishable business data were aggregated for analytics objectives oriented toward forecasting, demand forecasting, inventory optimization, loss prevention, downtime forecasting, and optimization of shipping sequences, among others. Business analytics-based Decision Support Systems mapping on Smart BDA Technologies have been developed in the manufacturing and supply-chain sector, with illustrations for manufacturing and shipping. Smart BDA Technologies are examined to evaluate the presumed positive contribution toward the innovation and operations of a business. By integrating deep learning for predictive analytics, the architecture is designed to accommodate Big Data analytics beyond the presentation layer within a big-data ecosystem base, enabling new decision-support capabilities that advance operation, operations excellence, management, and innovation in its target sectors and firms. A comparison with similar proposed architectures has demonstrated how the comparison criteria and performance results indicate a competitive advantage in the proposed architecture.

7. Conclusion

AI-driven big data architecture offers industry-focused decision support systems that respond to data-driven inquiries and support evidence-based conclusions. A comprehensive overview of the full stack—data lakes, AI/ML pipelines, model repositories, and decision-support use cases in sectors such as manufacturing—highlights critical elements, enables reproduction, and stimulates real-world implementation.

8. References

1. Awan, U., Shamim, S., Khan, Z., Zia, N. U., Shariq, S. M., & Khan, M. N. (2021). Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. Technological Forecasting and Social Change, 168, 120766.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

2. Ranjan, J., & Foropon, C. (2021). Big data analytics in building the competitive intelligence of organizations. International Journal of Information Management, 56, 102231.
3. Nandan, B. P., & Chitta, S. S. (2023). Machine Learning Driven Metrology and Defect Detection in Extreme Ultraviolet (EUV) Lithography: A Paradigm Shift in Semiconductor Manufacturing. Educational Administration: Theory and Practice, 29 (4), 4555–4568. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.
4. Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Information & Management, 58(3), 103434.
5. Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine learning for industrial applications: A comprehensive literature review. Expert Systems with Applications, 175, 114820.
6. Kalisetty, S. (2023). Big Data–Driven Cloud Collaboration Models for Enhancing Supplier–Retailer Synchronization in Mod-ern Manufacturing Supply Chains. Journal of Computational Analy-sis and Applications (JoCAAA), 31(4), 2188-2205.
7. Mikalef, P., Conboy, K., & Krogstie, J. (2021). Artificial intelligence as a service: The case for an AI capability ecosystem. Information Systems Frontiers, 23, 1–16.
8. Fosso Wamba, S., Queiroz, M. M., Trinchera, L., & Shi, C. (2021). Big data analytics and artificial intelligence in supply chain management: A systematic literature review and research agenda. International Journal of Production Research, 59, 1–25.
9. Mashetty, S. (2023). A Comparative Analysis of Patented Technologies Supporting Mortgage and Housing Finance. Available at SSRN 5249181. Mashetty, S. (2023). A Comparative Analysis of Patented Technologies Supporting Mortgage and Housing Finance. Available at SSRN 5249181.
10. Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2021). Big data and dynamic capabilities: A bibliometric analysis and systematic literature review. Management Decision, 59(8), 1979–2001.
11. Akter, S., Fosso Wamba, S., Gunasekaran, A., Dubey, R., & Childe, S. J. (2021). How to improve firm performance using big data analytics capability and business strategy alignment? International Journal of Production Economics, 182, 113–131.
12. Pamisetty, V. (2023). Leveraging artificial intelligence for strategic decision-making in tax administration and policy design. Available at SSRN.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

13. Dubey, R., Bryde, D. J., Foropon, C., Graham, G., Giannakis, M., & Mishra, D. (2021). The role of artificial intelligence and big data in supply chain management: A resource-based view. International Journal of Production Research, 59(23), 1–20.
14. Adusupalli, B. (2022). The Impact of Regulatory Technology (RegTech) on Corporate Compliance: A Study on Automation, AI, and Blockchain in Financial Reporting. Mathematical Statistician and Engineering Applications, 71 (4), 16696–16710.
15. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.
16. Mashetty, S. (2023). View of A Comparative Analysis of Patented Technologies Supporting Mortgage and Housing Finance. Educational Administration: Theory and Practice.
17. Awan, U., Khan, M. N., & Zhang, X. (2021). Big data analytics and organizational decision-making: A systematic perspective on data-driven enterprise transformation. Journal of Enterprise Information Management, 34, 1–20.
18. Mandala, G., Reddy, R., Nishanth, A., Yasmeen, Z., & Maguluri, K. K. (2023). Ai and ml in healthcare: redefining diagnostics, treatment, and personalized medicine. International Journal of Applied Engineering & Technology, 5(S6).
19. Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. Engineering Science and Technology, an International Journal, 29, 101021.
20. Dorneanu, B., Zhang, S., Ruan, H., Heshmat, M., Chen, R., Vassiliadis, V. S., & Arellano-Garcia, H. (2022). Big data and machine learning: A roadmap towards smart plants. Engineering Management Journal, 9, 623–639.
21. Meda, R., & Pamisetty, A. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice, 29 (4), 5215–5233.
22. Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: A novel IoT and machine learning based

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

decision support system for predictive maintenance in Industry 4.0. Journal of Intelligent Manufacturing, 34, 107–121.

23. Johnson, M., Albizri, A., Harfouche, A., & Fosso-Wamba, S. (2022). Integrating human knowledge into artificial intelligence for complex and ill-structured problems: Informed artificial intelligence. International Journal of Information Management, 64, 102479.
24. Chatterjee, S., Chaudhuri, R., Gupta, S., Sivarajah, U., & Bag, S. (2023). Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm. Technological Forecasting and Social Change, 196, 122824.
25. Recharla, M., & Chitta, S. AI-Enhanced Neuroimaging and Deep Learning-Based Early Diagnosis of Multiple Sclerosis and Alzheimer's.
26. Morales-Serazzi, M., González-Benito, Ó., & Martos-Partal, M. (2023). A new perspective of BDA and information quality from final users of information: A multiple study approach. International Journal of Information Management, 73, 102683.
27. Nakashololo, T. M., & Iyamu, T. (2023). Decision support model for big data analytics tools. South African Journal of Information Management, 25(1), a1678.
28. Ghosh, A. K., Fattahi, S., & Ura, S. (2023). Towards developing big data analytics for machining decision-making. Journal of Manufacturing and Materials Processing, 7(5), 159.
29. Pillai, V. (2023). Integrating AI-driven techniques in big data analytics: Enhancing decision-making in financial markets. International Journal of Engineering and Computer Science, 12(7), 25774–25787.
30. Paleti, S. (2023). Transforming Money Transfers and Financial Inclusion: The Impact of AI-Powered Risk Mitigation and Deep Learning-Based Fraud Prevention in Cross-Border Transactions. Available at SSRN, 5158588.
31. Papadopoulos, T., & Balta, M. E. (2022). Climate change and big data analytics: Challenges and opportunities. International Journal of Information Management, 63, 102451.
32. Muhammad, A., Yu, C. K., Qadir, A., Ahmed, W., Yousuf, Z., & Fan, G. (2022). Big data analytics capability as a major antecedent of firm innovation performance. Journal of Information Technology, 37(4), 1–18.
33. Chen, H., Chiang, R. H. L., & Storey, V. C. (2021). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 45(1), 1–20.
34. Pandugula, C., & Nampalli, R. C. R. Optimizing Retail Performance: Cloud-Enabled Big Data Strategies for Enhanced Consumer Insights.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 18

35. Hirsch, E., Hoher, S., & Huber, S. (2023). An OPC UA-based industrial big data architecture. Proceedings of the International Conference on Industrial Informatics, 1–8.
36. Morales, M., González, Ó., & Martos-Partal, M. (2023). Exploring artificial intelligence and big data scholarship in information systems: A citation, bibliographic coupling, and co-word analysis. International Journal of Information Management Data Insights, 3(2), 100185.