MANTECH Publications

Journal of Converging Technologies in Intelligent
Volume 1, Issue 2, May-August 2025

Big Data Analytics and Predictive Intelligence for Optimizing Complex Decision-Making Processes: A Structured Review of Methods, Architectures, and Governance Challenges

Ananya R. Verma[1], Karan M. Srivastava[2], Neha P. Tiwari[3]

ABSTRACT

Complex decision-making in modern enterprises, healthcare systems, and supply networks is characterised by high-dimensional data, uncertainty, competing objectives, and compressed response times, conditions under which intuition-led and descriptive-reporting approaches increasingly fail. Big data analytics and predictive intelligence, the combination of scalable data infrastructure, machine learning, and optimisation, promise to convert such data into timely and defensible decisions. This structured narrative review synthesises 25 peer-reviewed studies and landmark publications spanning 1997–2023 to examine how big data analytics supports complex decision-making across five dimensions: conceptual foundations, predictive and prescriptive methods, domain applications, governance and explainability, and deployment engineering. Works were purposively selected using explicit criteria, classified thematically, and appraised for methodological transparency. The synthesis shows that (i) tree-ensemble and deep sequence models dominate predictive tasks, yet accuracy rarely improves decisions unless coupled with prescriptive optimisation; (ii) value realisation depends more on organisational analytics capability than on technology; (iii) explainability and privacy-preserving learning are prerequisites for adoption in regulated domains; and (iv) deployment debt, data drift, and feedback loops threaten long-run reliability. A five-layer integrative framework linking data, processing, prediction, prescription, and governance is proposed, together with research priorities for decision-centric evaluation, causal reasoning, and sustainable analytics operations.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

KEYWORDS: *Big data analytics, Predictive intelligence, Decision-making, Machine learning, Prescriptive analytics, Explainable AI, Decision support systems, MLOps, Data governance, Structured review*

INTRODUCTION

Decisions in contemporary organisations are rarely isolated choices among well-defined alternatives. Hospital capacity planning, retail inventory allocation, credit-risk management, and logistics routing each involve interdependent variables, incomplete information, delayed feedback, and trade-offs among cost, risk, and service quality. Such decisions are complex because their outcomes emerge from nonlinear interactions that unaided human judgement cannot reliably anticipate [1, 2]. At the same time, digital systems now generate data at a scale that exceeds conventional database and statistical tooling. The defining characteristics of this data, namely volume, velocity, and variety, were articulated early in the 2000s [3] and later extended with veracity, variability, and value [4].

Big data analytics (BDA) denotes the infrastructure, algorithms, and organisational practices used to extract actionable knowledge from such data, while predictive intelligence denotes the use of machine learning and statistical models to forecast future states and to recommend or automate responses [2, 5]. Business intelligence and analytics has progressed from structured, database-centred reporting towards web-based and sensor-driven analytics [2], and data science has been framed explicitly as a support for data-driven decision-making [5].

The stakes are practical as well as academic. In clinical settings, a mis-specified risk model can misallocate scarce beds or delay intervention; in supply networks, a biased forecast propagates through procurement, production, and distribution as amplified inventory error. These examples show that analytics must be judged by the decisions it changes, and by the reliability with which it does so over time, rather than by laboratory benchmarks alone [6, 7].

Despite this promise, three problems persist. First, predictive accuracy is frequently mistaken for decision quality: an accurate forecast does not by itself identify the best action under constraints [8]. Second, analytic outputs are often opaque, which hinders trust and accountability in high-stakes settings [9, 10]. Third, models that perform well during development degrade in production because of data drift, entangled pipelines, and feedback loops [6]. The central problem addressed here is therefore how big data analytics and predictive

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

intelligence can be systematically designed, governed, and evaluated so that they optimise, rather than merely inform, complex decision-making. The paper contributes a structured synthesis of the field, a comparative assessment of predictive and prescriptive approaches, and an integrative five-layer framework.

LITERATURE REVIEW

The literature is organised into four strands: conceptual foundations and value creation, analytical methods, domain applications, and governance, explainability, and deployment.

Conceptual Foundations And Value Creation

Davenport and Harris argued that organisations competing on analytics treat data-driven insight as a strategic asset rather than a reporting by-product [1]. Chen et al. traced business intelligence and analytics through three phases, from relational database-centred systems to web-based text and network analytics and, most recently, mobile and sensor-enabled data [2]. Gandomi and Haider surveyed methods suited to structured and unstructured data, including text, audio, video, and social media analytics [4]. On value creation, Wamba et al. combined a systematic review with a longitudinal case study to identify the pathways through which big data generates business value [11], and Mikalef et al. synthesised the literature on analytics capabilities, framing capability, the combination of data, technology, skills, and routines, as the unit through which investments generate performance [12].

Predictive And Prescriptive Methods

Random forests established that ensembles of decorrelated trees generalise well and resist overfitting [13], and gradient-boosted trees, notably XGBoost, delivered scalable, sparsity-aware learning that became a default for tabular business data [14]. For temporal data, long short-term memory networks addressed the vanishing-gradient problem of recurrent architectures through gated memory cells [15]. Kraus et al. reviewed deep learning in business analytics and operations research, highlighting both its predictive potential and its managerial implications, including limited transparency [16]. Prediction alone, however, does not prescribe action. Bertsimas and Kallus proposed combining machine learning with optimisation so that covariate-informed predictions drive decisions under constraints, moving from predictive to prescriptive analytics [8]. Formally, given covariates x describing the decision context and an uncertain outcome Y, a prescriptive policy selects the action z that minimises the expected cost $E[c(z; Y) \mid X = x]$ over a feasible set Z, so that a conditional distribution learned from data replaces the fixed-parameter assumptions of classical

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

optimisation models [8]. Scalable execution is enabled by distributed engines such as Apache Spark, which unify batch, streaming, SQL, and machine-learning workloads [17].

Domain Applications

In healthcare, Raghupathi and Raghupathi described the promise of analytics for clinical decision support, population health, and cost control, and proposed an architectural framework for healthcare analytics [18], while Rajkomar et al. discussed the clinical potential of machine learning alongside risks from biased data and the need for regulatory oversight [19]. In supply chains, Waller and Fawcett argued that predictive analytics and big data would transform network design and management [20]. In retail, Fildes et al. documented advances in demand forecasting and the continuing gap between research methods and operational practice [7]. Taken together, these studies show that the same analytic pipeline recurs across domains while the binding constraints differ: clinical safety and bias in health care, network-wide coordination in supply chains, and forecast-to-order alignment in retail.

Governance, Explainability, And Deployment

Explainability research offers model-agnostic attribution through SHAP [9] and local surrogate explanations through LIME [21], while Doshi-Velez and Kim proposed a taxonomy for evaluating interpretability [10] and Gunning and Aha described programme-level efforts towards explainable AI [22]. Privacy constraints are addressed through differential privacy [23] and federated learning, which trains shared models without centralising raw data [24]. Sculley et al. warned that machine-learning systems accumulate hidden technical debt through entanglement, undeclared data dependencies, and feedback loops [6], and Kreuzberger et al. consolidated MLOps principles and architecture for operationalising models reliably [25].

RESEARCH GAP

Existing reviews typically address one stratum in isolation, whether infrastructure [17], algorithms [16], domain value [11, 18, 20], or interpretability [9, 10], and rarely examine how these strata interact to determine decision quality. Three gaps are evident. First, the transition from predictive to prescriptive analytics is discussed conceptually [8] but seldom linked to the data-engineering and governance preconditions for deployment. Second, explainability and privacy are treated as add-ons rather than design constraints of the analytic pipeline. Third, evaluation emphasises predictive metrics such as accuracy and error rather than decision-level outcomes. This review addresses these gaps by synthesising the strata within a single decision-centric framework.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

OBJECTIVES

- To trace the conceptual evolution of big data analytics and predictive intelligence as enablers of complex decision-making.
- To compare predictive and prescriptive modelling approaches in terms of strengths, constraints, and decision suitability.
- To synthesise evidence on applications in healthcare, supply chain, and retail decision contexts.
- To examine explainability, privacy, and deployment engineering as determinants of sustained decision quality.
- To propose an integrative framework and research agenda for decision-centric analytics.

METHODOLOGY

Review Design And Selection Protocol

A structured narrative review was conducted, applying the transparency principles of systematic review guidelines [26] and the reporting logic of the PRISMA 2020 statement [27]. The review is purposive rather than exhaustive: instead of enumerating all publications, it selects influential works that define each stratum of the analytics-to-decision pipeline. Candidate works were identified by citation chaining from foundational papers in business intelligence and machine learning and by inspecting the reference lists of recent surveys. Table 1 summarises the protocol.

Table 1. Review Protocol and Selection Criteria

Element	Specification
Review type	Purposive structured narrative review
Time span	1997–2023
Corpus	25 works: 17 journal articles or monographs, 5 conference papers, 1 preprint, 1 book, 1 research note
Inclusion	Foundational or influential; addresses analytics, prediction, prescription, governance, or deployment for decisions; English language
Exclusion	Vendor white papers; non-peer-reviewed commentary (except foundational notes); statistical theory without decision relevance
Appraisal and synthesis	Qualitative appraisal of grounding, transparency, and limitations; thematic coding; no meta-analysis

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Classification And Synthesis

Each of the 25 selected works was coded by theme (foundations, methods, applications, governance, deployment), publication period, and decision relevance. Thematic synthesis then compared claims, assumptions, and reported limitations across studies, and coded frequencies were used to describe the shape of the evidence base (Figure 1). Quality was appraised qualitatively through three questions: is the contribution theoretically grounded, is the method transparent, and are limitations acknowledged? Quantitative meta-analysis was not attempted because the studies use heterogeneous designs and outcome measures.

Throughout the synthesis, each work was examined through a decision-pipeline lens with five questions: what data it assumes, how predictions are produced, how predictions become actions, how trust is established, and how reliability is sustained. These questions correspond to the layers of the framework presented later, so that the framework is derived from, rather than imposed on, the evidence.

RESULTS AND FINDINGS

Figure 1 shows the distribution of the 25 reviewed works by theme and period. Applications and governance studies are the largest categories (six works each), followed by foundations and methods (five each) and deployment (three). Ten works (40%) appeared between 2015 and 2018, coinciding with the maturation of scalable machine-learning tooling, while all six governance works date from 2014 onwards, reflecting growing regulatory and societal scrutiny. Application studies span healthcare (two works), supply chains (one), retail (one), and cross-sector organisational value (two), indicating that evidence is richest where data are plentiful and outcomes are measurable.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

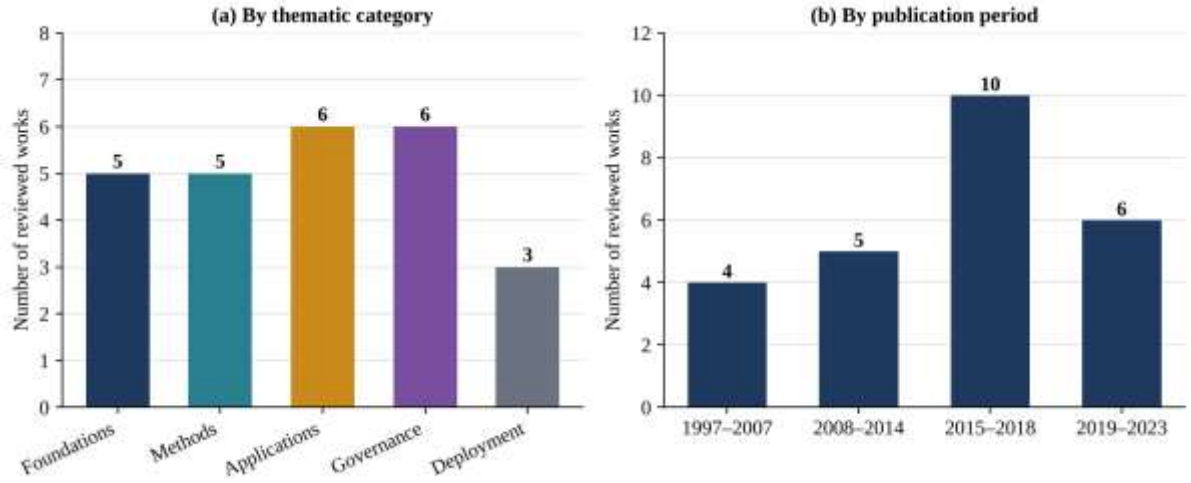


Figure 1. Distribution of the 25 reviewed works by (a) thematic category and (b) publication period

Finding 1: prediction is necessary but insufficient. Tree ensembles and deep sequence models provide the strongest predictive foundations for tabular and temporal data respectively (Table 2), yet only the formulation of Bertsimas and Kallus integrates predictions explicitly into constrained optimisation [8]. In the reviewed corpus, one of five methods studies addresses prescription directly, indicating that most method development stops at prediction.

Table 2. Comparison of Predictive and Prescriptive Approaches for Decision Support

Approach	Key works	Strengths	Constraints	Decision role
Tree ensembles	[13, 14]	Accurate on tabular data; scalable	Opaque without post-hoc tools	Risk scoring, prioritisation
LSTM and deep models	[15, 16]	Learn long temporal dependencies	Data-hungry; low transparency	Demand and patient-trajectory forecasting
Prescriptive ML with optimisation	[8]	Links predictions to constrained actions	Needs explicit objectives and constraints	Allocation, pricing, scheduling
Distributed platforms	[17]	Unify batch, stream, SQL, and ML workloads	Engineering overhead	Data preparation and serving at scale

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Finding 2: capability governs value. Application studies consistently report that benefits depend on integration with organisational processes, skills, and data quality [11, 12, 18, 20], and Fildes et al. document a persistent research–practice gap in retail forecasting [7]. Finding 3: governance is a precondition in regulated domains. Explainability tools [9, 21], privacy-preserving learning [23, 24], and attention to data bias [19] recur as enabling conditions for clinical and financial adoption. Finding 4: deployment debt erodes reliability. Hidden dependencies and feedback loops [6] and the need for continuous monitoring and automation [25] indicate that sustained decision quality is an engineering problem as much as a modelling problem. Table 3 consolidates these findings.

Table 3. Thematic Synthesis of Findings Across the Reviewed Works

Theme	Works (n)	Principal finding	Decision implication
Foundations	5	Data characteristics and BI evolution define analytic scope [1–5]	Match ambition to data maturity
Methods	5	Ensembles and sequence models lead prediction; prescription is rare	Evaluate on decision outcomes
Applications	6	Value depends on capability and domain integration	Embed analytics in workflows
Governance	6	Explainability and privacy enable adoption [21, 22]	Treat as design constraints
Deployment	3	Technical debt and drift erode reliability	Invest in MLOps monitoring

A cross-cutting observation is the tension between predictive performance and interpretability. The high-performing model families in Table 2 are the least transparent, whereas the explanation methods that mitigate this opacity [9, 21] are approximations that must themselves be validated [10]. Decision-makers therefore face a design choice among inherently interpretable models, post-hoc explanation of complex models, and human-in-the-loop review, with the appropriate choice depending on the cost of error and the regulatory context [22].

DISCUSSION

Integrative Decision-Centric Framework

The synthesis yields the five-layer framework in Figure 2. The data layer collects heterogeneous sources; the processing layer applies distributed engines [17]; the prediction layer deploys ensemble and sequence models [14, 15]; the prescription layer converts

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

predictions into constrained actions [8]; and the decision-support layer presents recommendations for human review or automated execution. Governance is a cross-cutting layer covering explainability, privacy, and monitoring [9, 24, 25]. A feedback loop returns observed outcomes to the data layer and must be monitored, because decisions influence future training data [6].

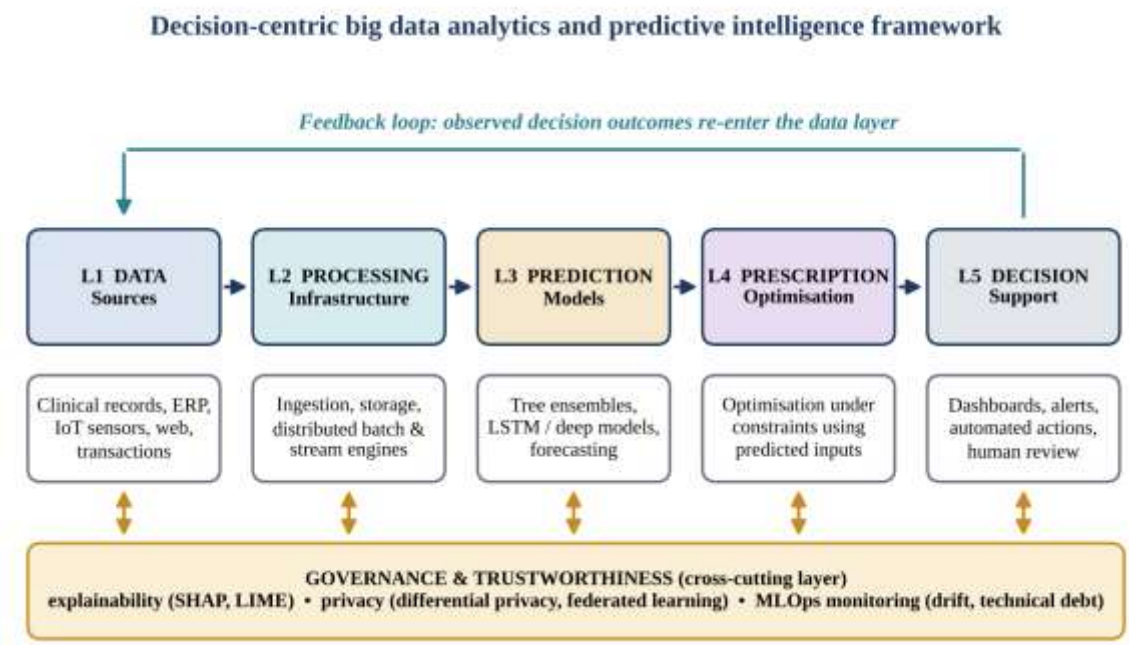


Figure 2. Proposed five-layer decision-centric framework for big data analytics and predictive intelligence with cross-cutting governance and feedback

Figure 3 situates these layers on an analytics maturity continuum, from descriptive and diagnostic analytics to predictive and prescriptive analytics. Each step addresses greater decision complexity but also increases automation and governance burden, which explains why organisations that skip foundational data and governance work struggle to realise prescriptive value. The continuum also clarifies evaluation: descriptive and diagnostic stages are judged by data accuracy and timeliness, predictive stages by calibration and error, and prescriptive stages by realised cost, risk, and service outcomes, which is why decision-level metrics become indispensable at the highest maturity stage.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

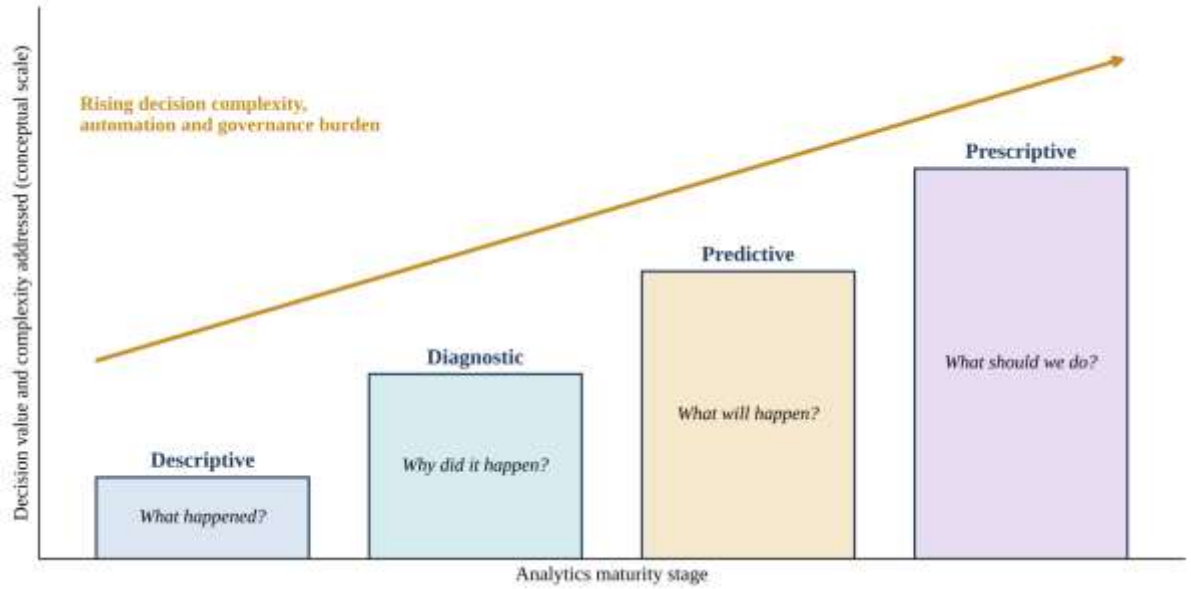


Figure 3. Analytics maturity continuum from descriptive to prescriptive analytics (conceptual illustration)

Scientific And Engineering Significance

The review reframes the success criterion of analytics from model accuracy to decision quality, aligning machine-learning practice with operations research [8, 16]. It also clarifies that interpretability, privacy, and operational reliability are design constraints interacting with predictive performance, rather than post-hoc additions [6, 10]. A further contribution is methodological: by treating each layer as a testable component, the framework supports ablation-style evaluation in which the contribution of prediction, prescription, and governance to decision quality can be isolated empirically, a design largely absent from the reviewed work.

Practical Applications

In healthcare, the framework supports capacity planning and clinical risk stratification under regulatory oversight [18, 19]. In supply chains and retail, it links demand forecasts to inventory and network decisions [7, 20]. Practitioners should begin with explicit decision mapping, define objectives and constraints before selecting models, embed explanation and privacy controls from the outset, and monitor drift continuously after deployment. Implementation should proceed incrementally: a pilot can target a single high-value decision, establish a baseline from current practice, compare model-assisted decisions against that baseline on realised outcomes, and extend the pipeline only when monitoring shows stable performance. Staged adoption limits the exposure created by hidden dependencies and feedback loops [6, 25] and gives governance teams time to establish audit trails.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

LIMITATIONS

- The review is purposive rather than exhaustive; selection bias is possible, and the corpus ends in 2023, so recent developments such as foundation-model-based decision support are not covered. Restriction to English-language sources may also omit regional evidence from emerging economies.
- No quantitative meta-analysis was possible because reviewed studies use heterogeneous designs, domains, and outcome measures; synthesis is therefore qualitative, and effect sizes or confidence intervals could not be pooled across studies.
- Reported results in the reviewed studies come from different data sets and settings, so comparisons in Table 2 reflect qualitative characteristics rather than head-to-head performance.
- Thematic coding and quality appraisal were performed by the authors without formal inter-rater reliability assessment, and Figure 3 is a conceptual illustration rather than an empirical result.

FUTURE SCOPE

- Causal and counterfactual analytics that estimate the effect of interventions, not only future outcomes, to support robust prescriptive decisions.
- Decision-level benchmarks and evaluation protocols that measure realised cost, risk, and service outcomes in addition to predictive accuracy.
- Governance-aware evaluation of large language and foundation models as decision-support components, including auditability, privacy, and human oversight.
- Empirical studies of human–AI collaboration testing how explanation formats and decision rights affect the acceptance, calibration, and accountability of analytic recommendations.
- Energy-efficient and sustainable analytics pipelines, and federated architectures for cross-organisational decision-making without raw data sharing.

CONCLUSION

Big data analytics and predictive intelligence can optimise complex decision-making only when prediction is coupled with prescription, supported by organisational capability, and constrained by explainability, privacy, and operational discipline. This review of 25 works demonstrates that the field's methods are mature while its decision-level evaluation and governance practices lag behind. The proposed five-layer framework and research agenda provide a basis for designing analytics systems judged by the quality of the decisions they

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

enable. For practitioners, the implication is to invest first in decision definition, data quality, and governance, and only then in model sophistication.

REFERENCES

1. Davenport TH, Harris JG. *Competing on Analytics: The New Science of Winning.* Harvard Business School Press, Boston. 2007.
2. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: From big data to big impact. *MIS Quarterly.* 2012; 36(4): 1165–1188.
3. Laney D. 3D data management: Controlling data volume, velocity, and variety. *META Group Research Note.* 2001; 6(70): 1.
4. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management.* 2015; 35(2): 137–144.
5. Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. *Big Data.* 2013; 1(1): 51–59.
6. Sculley D, Holt G, Golovin D, et al. Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems.* 2015; 28: 2503–2511.
7. Fildes R, Ma S, Kolassa S. Retail forecasting: Research and practice. *International Journal of Forecasting.* 2022; 38(4): 1283–1318.
8. Bertsimas D, Kallus N. From predictive to prescriptive analytics. *Management Science.* 2020; 66(3): 1025–1044.
9. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems.* 2017; 30: 4765–4774.
10. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608.* 2017.
11. Wamba SF, Akter S, Edwards A, Chopin G, Gnanzou D. How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics.* 2015; 165: 234–246.
12. Mikalef P, Pappas IO, Krogstie J, Giannakos M. Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management.* 2018; 16(3): 547–578.
13. Breiman L. Random forests. *Machine Learning.* 2001; 45(1): 5–32.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

14. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016; 785–794.
15. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation.* 1997; 9(8): 1735–1780.
16. Kraus M, Feuerriegel S, Oztekin A. Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research.* 2020; 281(3): 628–641.
17. Zaharia M, Xin RS, Wendell P, et al. Apache Spark: A unified engine for big data processing. *Communications of the ACM.* 2016; 59(11): 56–65.
18. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems.* 2014; 2: 3.
19. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New England Journal of Medicine.* 2019; 380(14): 1347–1358.
20. Waller MA, Fawcett SE. Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics.* 2013; 34(2): 77–84.
21. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016; 1135–1144.
22. Gunning D, Aha DW. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine.* 2019; 40(2): 44–58.
23. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science.* 2014; 9(3–4): 211–407.
24. McMahan HB, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (PMLR).* 2017; 54: 1273–1282.
25. Kreuzberger D, Kühl N, Hirschl S. Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE Access.* 2023; 11: 31866–31879.
26. Kitchenham B, Charters S. *Guidelines for performing Systematic Literature Reviews in Software Engineering (EBSE Technical Report EBSE-2007-01).* Keele University and University of Durham. 2007.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved

27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021; 372: n71.

***Author for Correspondence:**
Ananya R. Verma
E-mail: ananya.verma.cse897@gmail.com

[1]Associate Professor, Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh
[2]Assistant Professor, Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh
[3]Research Scholar, Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh

Received Date: May 02, 2025
Accepted Date: May 15, 2025
Published Date: May 29, 2025

Citation: Ananya R. Verma, Karan M. Srivastava, Neha P. Tiwari. Big Data Analytics and Predictive Intelligence for Optimizing Complex Decision-Making Processes: A Structured Review of Methods, Architectures, and Governance Challenges. Journal of Converging Technologies in Intelligent. 2026; 1(2): 75-88p.

Page 75-88 © MANTECH PUBLICATIONS 2025. All Rights Reserved