SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

BIG DATA ANALYTICS FOR PREDICTIVE DECISION MAKING IN COMPLEX BUSINESS ENVIRONMENTS

Sheharyar Hussain[*1], Rana Muhammad Yasir Naseem[2]

[*1]*Department of Computing, Riphah International University, Pakistan*
[2]*Department of Mechanical, Mechatronics and Manufacturing Engineering, University of Engineering and Technology (UET), Pakistan*

[*1]sheharyarhussain23@gmail.com, [2]yasirrajpoot570@gmail.com

DOI: https://doi.org/10.5281/zenodo.21187418

Keywords
Big Data Analytics, Predictive Analytics, Decision Support Systems, Machine Learning, Business Intelligence, Data-Driven Decision Making, Organizational Performance, Implementation Challenges

Article History
Received: 25 April 2026
Accepted: 04 June 2026
Published: 21 June 2026

Corresponding Author: *
Sheharyar Hussain

Abstract

Big Data Analytics (BDA) has evolved as a strategic capability that allows enterprises to deal with the ever-increasing complexity and volatility of the business environment. This survey review comprehensively evaluates the impact of predictive analytics in decision-making across a variety of contexts within an organization and synthesized evidence from 109 academic and practitioner journal articles, published between 2010 and 2024. Through our extensive literature analysis, we build a systematic taxonomy of the Big Data Predictive Analytics (BDPA) applications within diverse organizational domains: industrial manufacturing, healthcare, financial services, e-commerce, strategic management, smart agriculture, and the information and communication technology (ICT) sector. This work scrutinizes the key theoretical aspects of predictive analytics, its location in the continuum from descriptive to prescriptive Analytics, and systematically analyzes various analytical techniques used for making predictions, ranging from the traditional statistical approaches to sophisticated machine learning algorithms. Key insights from our review show that those organizations that have achieved advanced Analytics maturity are well-equipped to reap several substantial benefits: reduction of unplanned downtime (20%-40%), enrichment of customer knowledge, enhancement of the financial health of the organization, and the gaining of sustainable competitive advantage. However, we acknowledge the numerous persistent challenges to implementation, such as problems with the quality of the data, deficiencies in required skills, ethical implications of the use of predictive analytics, and the resistance within the organization to the adoption of data-driven decision making. We offer an Integrated conceptual framework connecting BDA with decision processes and business outcomes, highlight key facilitators and inhibitors to effective implementation. We investigate various novel technologies like Explainable Artificial Intelligence (XAI), Edge Analytics, and Privacy-Preserving Analytics that hold the promise to overcome existing limitations. In addition to bridging the gap between academic scholarship and practitioner experience in the application of BDPA, the survey review suggests the need to develop explainable predictive models, audit for fairness, and create models for ethical AI implementation as key future research directions.

1. Introduction

1.1 Background and Context

Current organizations functioning under a highly complex, volatile, uncertain, and data-intensive environment are considered as the new business landscape (Chaudhary et al., 2024). The advent and ubiquity of digital technologies such as those of the Internet of Things (IoT) sensors, social media channels, mobile devices, and organizational system have produced immense, varied, and ever-streaming data that has changed how organizations perceive their activities, markets, and various stakeholders. In an increasingly complex world, this information deluge offers huge opportunities to decision-making personnel, accompanied by immense challenges (Benbya et al., 2020).

Traditional ways of decision-making which were primarily dependent on managerial instincts, prior references, and superficial statistical analysis are falling far short in the face of uncertainty and dynamism of the contemporary business environment.

While data-rich yet wisdom-starved, contemporary organizations face an information surplus – paradox without comprehension; and this wide divergence between data capacity and decision capability have recently spurred much curiosity and emphasis on the use of big data analytics (BDA) for revolutionizing the business intelligence system (Mikalef et al., 2019). BDA can be defined as the systematic process of applying various advanced analytical approaches, for instance, those relating to machine learning, statistic modeling, data mining and artificially intelligence technologies on large volume, wide variation and higher-velocity data in order to mine the hidden knowledge, foresee the possible results and create proactive information so as to assist decision making (Jamarani et al., 2024). Predictive analytics as one of the highly advanced fields within BDA is designed to discover unknown future from the already existing data. Using machine learning, data mining, AI, statistics and related advanced analytic disciplines and technique, predictive analytics can foresee many aspects, such as the customer's preferences, trends of marketing, potential risks and strategic chances more or less accurately.

2. Theoretical Foundations and Conceptual Background

2.1 Defining Big Data and Its Characteristics

The term "big data" has transformed from merely a technical definition of a huge amount of data to a broader framework consisting of capabilities, technologies, and practices (Demchenko et al., 2014). Among others, the most common definition of big data highlights the following five "Vs" of the big data concept: Volume refers to data in enormous amounts that usually span across petabytes or terabytes. Big data technologies facilitate capturing more data in higher resolution at a faster pace than ever before because more devices connect through the network (e.g. Io Ts), many social behaviors are logged into the web, and transaction and device logs expand rapidly in scope (Shin & Choi 2025).

Volume denotes the amounts of data being generated or processed on a daily basis. A typical smart city system may need to handle big data in size of hundreds to thousands to petabytes, and a complex big data processing and managing framework for smarter societies require an in-depth understanding for the data structures involved, along with algorithms designed to optimize resource usage. Volume of data that needs to be managed today typically amounts to terabytes of data. It's already very common for a business to manage volumes of data in petabytes.

Digital transformation has produced many of data such as sensors or machine to machine that grows up with an incredible speed of growth in recent days. Big data, or massive and complex data that are difficult to process through conventional data management tools, is increasing with a growing pace. Velocity measures speed of data generated and analyzed on a real-time basis or near real-time. The data should be captured and processed so quickly that insights are gained immediately to support decisions made (Mikalef, 2018).

Variety implies diversity in the type of data generated and captured such as structured data (numerical and quantitative values), unstructured data (text and speech) and semi structured data (xml files) which require diverse tools for efficient management (Nwaimo, 2019). The range, diversity and richness of data type will create value from big data with analytics that exploit the variety such as audio and visual, text and speech (including mixed data of natural language). Variety describes variety in data type and content and they can be both structured, semi-structured, and unstructured.

The complexity in format means that organizations and IT practitioners must incorporate a variety of analytical techniques and tools in order to make sense out of them.

Variety refers to multiple data types such as textual data, audio data, video data, image data, web log data etc. This is also including multiple data sources which produce data. Data can be categorized into structured data, semi-structured data and unstructured data. Huge amounts of mixed data, such as audio data, visual data, text data, can be generated.

Veracity indicates uncertainty in the data quality. The data should be managed by quality processes to minimize ambiguity, inconsistency, and incompleteness of data (Borgman, 2017). It is very crucial to ensure data validity for obtaining meaningful knowledge. In a real-world Big Data scenario (as shown in figure 1), it can be assumed that some data might be of questionable veracity and should be identified.

Valuetraditionally defined Big Data in 1990's has expanded the scope of its original, limited scope of a specific technology towards Big Data capacity, capability or strategy. Value reflects the real goal: generating business outcomes from Big Data analytics (Abdalla, 2022).

Table 1: The Five Dimensions of Big Data Analytics

Dimension	Description	Business Implications
Volume	Scale of data generated and stored	Requires scalable storage and processing infrastructure
Velocity	Speed of data generation and processing	Demands real-time or near-real-time analytical capabilities
Variety	Diversity of data types and sources	Necessitates flexible integration and analysis approaches
Veracity	Quality, accuracy, and reliability of data	Requires robust data governance and quality management
Value	Worth derived from data through analysis	Drives investment justification and outcome measurement

2.2 The Analytics Continuum: Descriptive, Predictive, and Prescriptive

Big Data analytics, viewed as a journey, is imagined as a pathway to increasingly advanced analysis and decision support functions as businesses evolve from descriptive to predictive and prescriptive analytics. The least sophisticated descriptive analytics is aimed at answering the question, "What happened?" through aggregation of historical data into dashboards, performance dashboards, reports and key performance indicators that reflect past activities and present status but offer little future predictive insight (Bari & Ara, 2024).

Next, predictive analytics builds on this, answering, "What is likely to happen?"
by developing predictive models of expected outcomes using a variety of statistical models, including machine learning and forecast methods. This is typically applied to activities that may involve anticipating customer churn, future product and service demands, or the probability of credit risk or equipment maintenance needs (Shafa, 2025). The highest analytical maturity-prescriptive analytics-seeks to provide the answer, "What should we do?" by defining courses of action for business problems.

This is achieved by optimizing the choices based on a range of alternative possible courses of action and likely outcomes as defined by predictive models through techniques such as simulation and decision modeling, which lead directly to actionable, optimal decision-making (Demirbaga et al., 2024).
Existing literature and empirical studies indicates that business outcomes increase across the analytical continuum with both descriptive and prescriptive analytics achieving better business results, although progress from one stage to another requires considerable commitment to technology, skills, and change management (Demirbaga et al., 2024).

2.3 Decision-Making in Complex Business Environments

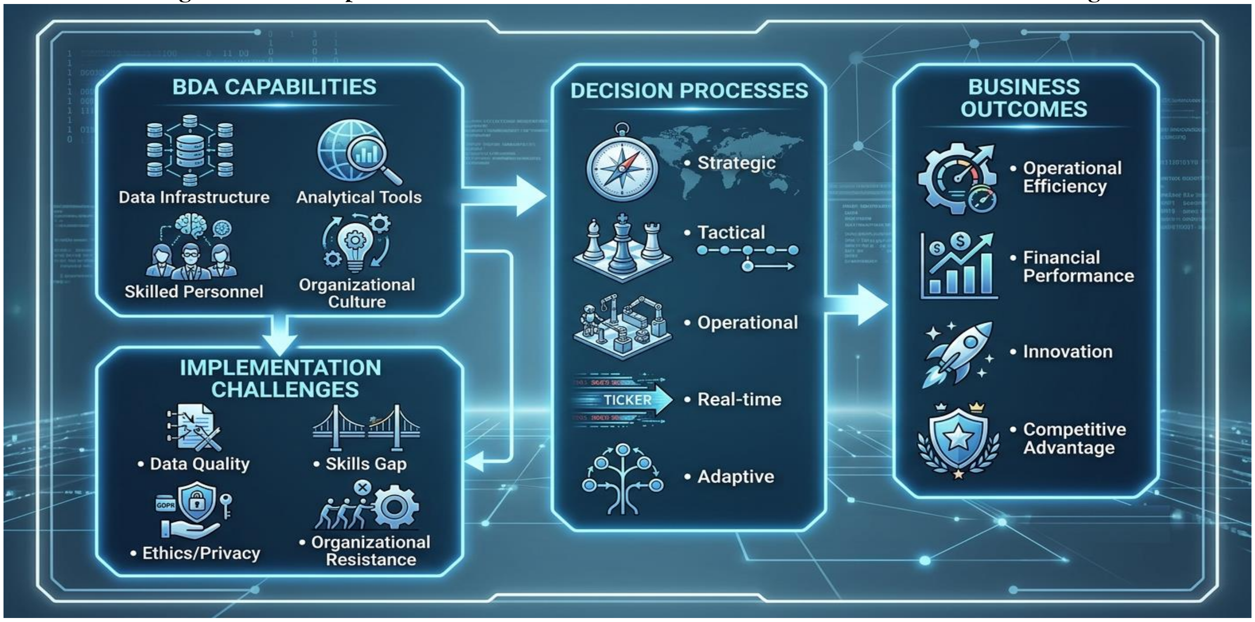
Complex business environments exhibit characteristics that challenge traditional decision-making approaches:
Uncertainty arises from incomplete information, rapid change, and unpredictable external factors. Decision-makers cannot fully anticipate future conditions, making forecasting essential yet inherently limited (Snowden & Boone, 2007).

Interconnectedness means decisions in one domain affect others through complex causal pathways. Supply chain disruptions, for example, cascade through production, logistics, finance, and customer relationships.

Information Asymmetry exists when different stakeholders possess different knowledge and perspectives. Organizations must integrate diverse information sources while recognizing their limitations (Noyes et al., 2012).

Time Pressure demands rapid decisions in increasingly dynamic markets. Delays in recognizing emerging threats or opportunities can have significant competitive consequences.
Ambiguity arises when data supports multiple interpretations. The same information may signal different implications depending on analytical frameworks and assumptions.
Big Data Analytics addresses these challenges by enabling evidence-based decision-making, reducing reliance on intuition, and supporting systematic analysis of complex phenomena. However, effective integration requires careful attention to the relationship between analytical outputs and human judgment (Fabac, 2010).

Figure 1: Conceptual Framework for BDA-Enabled Predictive Decision-Making



According to the framework, there are three overlapping constructs :

1. BDA capabilities: This refers to technical, people and organizational resources and capacities for doing analysis, which consists of data infra-structure, analytical tools, trained personnel and enabling cultural environment.

2. Decision making processes: These are organi-zational ways of decision making by utilizing analy-sis results, which ranges from strategic thinking to operative actions in real time.

3. Business outcomes: These are business per - formance consequences that result from BDA-enhanced decision making such as, financial benefits, efficiencies, innovations and competitive advantages. The decision to process to outcome translation is influenced by implementation challenges which represent obstacle that prevent firms from success fully adopting BDA (Pathak et al., 2023)

3. Taxonomy of Big Data Predictive Analytics Applications

3.1 Overview of Application Domains

This Section will give you a complete typology of BDPA applications by analyzing in a rigorous manner the 109 articles that we read through. Adopting the typology of Jamarani et al. (2024), the identified types are classified in 7 broader application sectors, each having their own subdivisions in terms of the types of business practices it encompasses.

Table 2: Taxonomy of Big Data Predictive Analytics Applications

Domain	Business Subcategories	Key Predictive Applications
Industrial	Manufacturing, Supply Chain, Production	Demand forecasting, predictive maintenance, quality prediction, logistics optimization
Financial Services	Banking, Insurance, Investment	Credit scoring, fraud detection, algorithmic trading, risk assessment
E-Commerce & Retail	Online Retail, Physical Retail	Customer lifetime value prediction, recommendation systems, churn prediction, dynamic pricing
Healthcare	Healthcare Administration, Services	Patient readmission prediction, resource planning, treatment outcome prediction
Strategic Management	Corporate Strategy, Operations	Market trend analysis, competitive intelligence, risk forecasting
Smart Agriculture	Farming, Agri-business	Crop yield prediction, precision agriculture, weather impact analysis
ICT & Technology	Telecommunications, Technology	Network performance prediction, capacity planning, user behavior forecasting

3.2 Industrial and Manufacturing Applications

Using Big Data & Predictive Analytics (BDPA), manufacturing organizations are revolutionizing their factories from dumb machinery centres into smart, intelligent and agile living operations across the entire value chain from the factory floor all the way up to the customer: Predictive Maintenance : Leveraging real time sensor data to predict when equipment is likely to fail and scheduling appropriate maintenance in time reducing unexpected downtime (Gomaa, 2025). Demand forecasting: Forecasting the actual customer demand using sales data coupled with marketing information. Supply Chain optimisation : integrating isolated logistics information from different supply chain partners to predict disruption &rerouting material. Quality prediction: Employing various models for to predict quality defect in to avoid product wastage. Together BDPA applications allow an organisation to shift from a responsive or reactive approach toward to proactive, efficient & reliable manufacturing operation, resulting in much lower cost, higher profit margin, & high-quality product (Jamarani et al., 2024).

3.3 Healthcare and Pharmaceutical Applications

Healthcare organizations face complex decisions involving patient outcomes, resource allocation, and cost management. BDPA offers significant potential for improving healthcare delivery. Key applications include:

Patient Readmission Prediction identifies patients at high risk of hospital readmission after discharge, enabling targeted follow-up care and interventions.

Resource Planning predicts patient volumes, staff requirements, and equipment needs, optimizing healthcare operations.

Treatment Outcome Prediction estimates likely patient responses to alternative treatments, supporting personalized medicine and evidence-based care (Ali, 2022).

4. Analytical Techniques and Technical Frameworks

4.1 Overview of Predictive Analytics Techniques

We can see that the papers researched present a good variety of analytical models chosen for predictive applications, also having a variation regarding the characteristics of these applications, available data and the analytic targets (Table 2).

Table 3: Predictive Analytics Techniques and Applications

Technique	Key Characteristics	Common Business Applications	Strengths	Limitations
Regression Analysis	Statistical modeling of relationships between variables	Demand forecasting, pricing, risk assessment	Interpretable, well-understood, efficient	Assumes linear relationships, sensitive to outliers
Time-Series Forecasting	Analysis of temporal patterns and trends	Sales forecasting, stock market prediction, resource planning	Captures temporal dynamics, seasonal patterns	Requires stationary data, limited for complex patterns
Decision Trees	Hierarchical decision rules based on data features	Credit scoring, customer segmentation, churn prediction	Highly interpretable, handles mixed data types	Prone to overfitting, unstable with small changes
Random Forests	Ensemble of decision trees with randomization	Fraud detection, risk prediction, recommendation	Robust to noise, handles high dimensionality	Less interpretable, computationally intensive
Support Vector Machines	Classification through optimal hyperplane separation	Classification tasks, anomaly detection, pattern recognition	Effective in high-dimensional spaces, memory efficient	Binary classification focus, sensitive to kernel selection
Neural Networks	Layered learning architectures inspired by neural systems	Image recognition, natural language processing, complex pattern prediction	Captures complex non-linear relationships, high accuracy	Requires large data, limited interpretability, computationally intensive

Deep Learning	Multi-layer neural networks with representation learning	Advanced prediction, voice recognition, autonomous systems	State-of-the-art accuracy on complex tasks	Massive data requirements, "black box" concerns
Gradient Boosting	Sequential addition of weak learners to optimize performance	Risk modeling, ranking, web search	High accuracy, handles missing data, robust	Computationally intensive, can overfit
Bayesian Methods	Probabilistic modeling with prior beliefs	Uncertainty quantification, Bayesian networks, decision support	Provides uncertainty estimates, incorporates prior knowledge	Computationally expensive for complex models

4.2 Machine Learning Approaches

Machine learning has emerged as the dominant paradigm for BDPA due to its ability to identify patterns in complex, high-dimensional data without explicit programming (Hasanuddin, 2025). Key approaches include:

Supervised Learning trains models on labeled historical data where the target outcome is known. The model learns to map inputs to outputs, enabling prediction when labels are unavailable. Common supervised techniques include regression, classification, and time-series forecasting.

Unsupervised Learning identifies patterns in unlabeled data, discovering natural groupings or associations. While less directly predictive, unsupervised approaches support exploratory analysis and feature discovery that enhances predictive models.

Reinforcement Learning learns optimal actions through trial and error, receiving feedback from system responses. This approach is particularly relevant for sequential decision problems and dynamic optimization.

Research demonstrates that machine learning models can achieve remarkable predictive accuracy, with healthcare applications reaching 99% diagnostic accuracy in specific contexts. However, model performance depends critically on data quality, feature engineering, and appropriate algorithm selection (Ponte, 2026).

4.3 Mathematical Formulation of the Prediction Problem

Let the dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^p$ is the feature vector and y_i is the target variable (Liu & Wu, 2007). For regression problems $y_i \in \mathbb{R}$, and for classification $y_i \in \{1,2,\ldots,K\}$.

The general predictive model is:

$$\hat{y}_i = f(x_i; \theta)$$

Where θ represents model parameters. The learning objective minimizes the empirical risk:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i; \theta)) + \lambda \Omega(\theta)$$

Where:

- L = Loss function
- Ω = Regularization term
- λ = Regularization parameter controlling complexity

4.4 Evaluation Metrics for Predictive Models

Assessing predictive model performance requires appropriate metrics aligned with business objectives (Manzoor et al., 2026). Common metrics include:

Accuracy represents the proportion of correct predictions among total predictions. While intuitive, accuracy can be misleading for imbalanced datasets where rare events are of primary interest.

Precision and Recall Offer better nuanced performance evaluation on classification tasks. Precision is about how many positive predictions were actually positive, whereas recall is about how many true positives were caught. These metrics can be very useful in scenarios like fraud detection where negative vs negative consequences are much more serious and have business impact.

F1 Score combines precision and recall as their harmonic mean, providing a single metric balancing both concerns.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) assesses the model's ability to distinguish between classes across threshold settings, providing a robust measure of classification performance (Islam & Ikbal, 2022).

Root Mean Square Error (RMSE) and **Mean Absolute Error (MAE)** are standard metrics for regression tasks, measuring average prediction error.

Business Metrics such as return on investment, cost savings, or revenue impact provide the ultimate performance measure, linking analytical models to organizational outcomes (Michael, 2025).

4.5 Lasso Regression (L1 Regularization)

Performs feature selection through sparse solutions:

$$\hat{\beta}^{Lasso} = \arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\}$$

This is equivalent to:
The Lasso regression problem is formulated as:

$$\min_{\beta} \| y - X\beta \|_2^2 \text{ subject to } \| \beta \|_1 \leq t$$

For classification tasks, logistic regression models the probability of binary outcomes as:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}} = \frac{1}{1 + e^{-X\beta}}$$

The **logit transformation**:

$$\log\left(\frac{P(Y = 1 \mid X)}{P(Y = 0 \mid X)}\right) = X\beta$$

Maximum likelihood estimation:

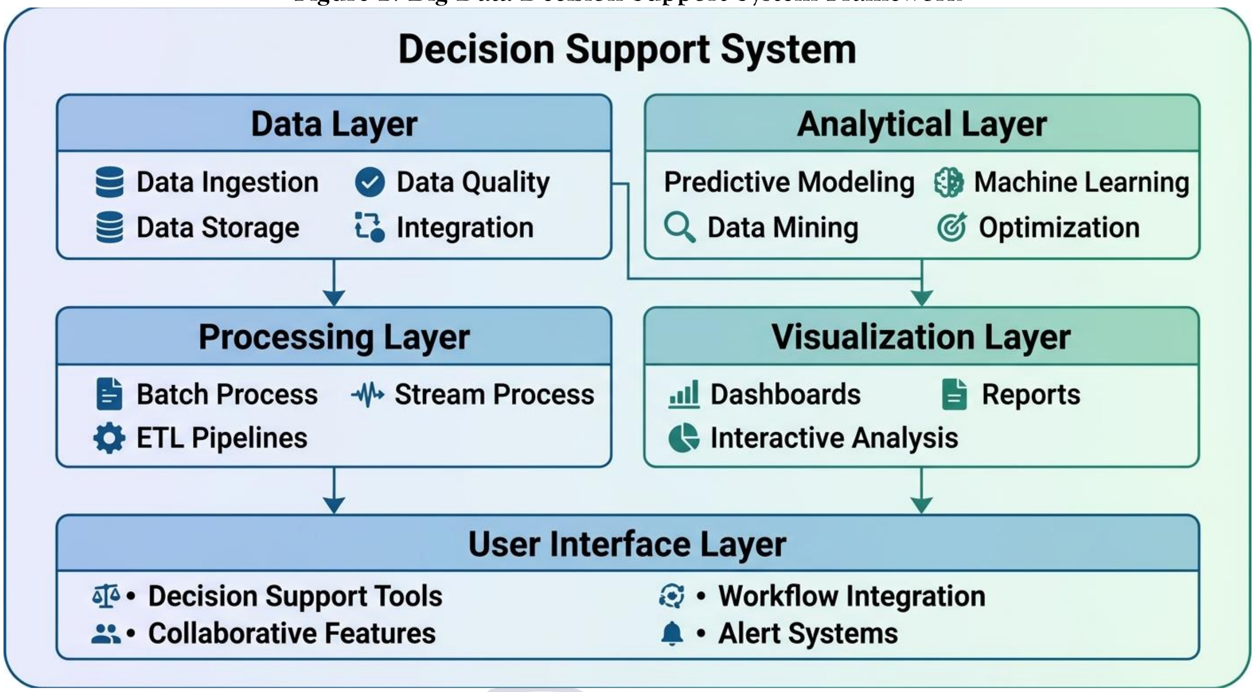
$$\hat{\beta} = \arg\max_{\beta}\sum_{i=1}^{n}[y_i \log(P_i) + (1 - y_i)\log(1 - P_i)]$$

Where $P_i = P(Y_i = 1 \mid X_i)$ (Glang et al., 2023).

4.6 Decision Support System Frameworks

Big Data Analytics for predictive decision-making systems has been typically achieved by Decision Support Systems (DSS), composed by the integration of the Data Processing components, Analytical Modeling, and User Interface, by means of software applications (Moreira et al., 2019).

Figure 2: Big Data Decision Support System Framework



High-performance DSS structures incorporate diverse application stages. These levels accommodate an organization's complete analytics cycle, beginning with data accumulation, storage and quality validation (data layer), through data change to its analytical presentation (processing layer); from predictive modelling and insight derivation (analytical layer) to their presentment through versatile interfaces (visualization layer); as well as decision-enabling features with interactive capacities and collaboration mechanisms for decision-makers (user layer) (Sauter, 2014).

Considerations concerning significant DSS features arescalability to process volumes of emerging data and flexible capability to manage different kinds of data and analytic techniques, usability for decision-makers in order to manage their analytical capacities, and integration capacity with business processes (Delen & Demirkan, 2013).

5. Implementation Challenges and Critical Success Factors

5.1 Data Quality and Governance

Data quality issues repeatedly emerge as the chief hurdle facing Big Data and Predictive Analytics (BDPA). There are various data quality dimensions where these issues show up. Missing data points tend to affect the reliability of analytics model and may also lead to potential bias, thus necessitate decisions like deleting or imputing missing values with potential implications for analytical findings (Choudhury & Pal, 2021). Incompatibilities between the formats, definitions, and standards of different data sources pose a data consistency problem, making data integration a complex issue and a solid Master Data Management and Data Governance plan to create a Single Source of Truth needed (Sodero et al., 2019). Another challenge can be data accuracy because some errors generated during capturing, processing or reporting that the analytical data processing pipelines inherit to challenge its validity (Bonthu, 2025). Timeliness is of critical relevance for most fast-changing business worlds, for instance; obsolete information will also bring us predictive relevance problems but performing data processing in real-time requires a very big infrastructural cost (Hikmawati, 2021). The ever-present challenge of data integration continues as organizations strive to connect varied internal and external sources, each differing in format, meaning, and quality (2023). It is through the removal of all these various barriers to quality-whether at a governing, technological, or organizational level-that effective BDPA outputs are able to materialize (Belani, 2025).

Table 4: Data Quality Challenges and Mitigation Strategies

Challenge	Description	Mitigation Approaches
Data Completeness	Missing or partial data records	Imputation methods, data augmentation, robust modeling
Data Consistency	Format and definition inconsistencies	Data governance, master data management, standardization
Data Accuracy	Errors in data collection and processing	Quality validation, error detection algorithms, data auditing
Data Timeliness	Outdated or stale data	Real-time processing, streaming analytics, periodic refresh
Data Integration	Heterogeneous data sources	ETL pipelines, data lakes, semantic integration
Data Provenance	Unclear data origins and transformations	Data lineage tracking, metadata management

5.2 Ethical, Privacy, and Regulatory Concerns

There are several critical ethical issues that businesses must address when deploying predictive analytics, namely privacy, fairness, accountability, transparency, and regulatory compliance. The deployment of technologies to collect and analyze personal information, especially sensitive personal data, necessitates the implementation of strong consent policies, disclosure mechanisms, and effective information safeguards to secure individual privacy (Masache & Tanner, 2026). There is also the problem that predictive algorithms tend to carry or magnify the same biases found in historic data, which can lead to and perpetuate unequal, discriminative outputs that may emerge in diverse social and demographic characteristics, especially for vulnerable population segments (Basu & Das, 2025). As it is known, the issue of establishing a person or entity who is responsible for the impact or output derived from algorithms used to infer or make recommendations, it can be extremely complicated and defy our understanding of legal and professional accountability and responsibility, not to mention how "black box" algorithms contribute to opacity which further hinder people's comprehension and the possibility of consenting to be subject to those processes, as well as makes auditing difficult (Sheng et al., 2025). These can be challenging in practice since they often clash with regulatory requirements emanating from laws like the GDPR or the CCPA which mandate data protection measures; and this reality compels organisations to consider and develop ethicalAI framework(s) but they may struggle to transform into audit able and enforcable corporate governance structures that may cope with evolving circumstances (Leon, 2026).

6. Business Benefits and Organizational Outcomes

6.1 Operational Efficiency Improvements

The extensive and pervasive impacts of Big Data and Predictive Analytics (BDPA) on both operational excellence and financial results are significant, and a formidable enabler of enterprise value creation. On the operational front, the predictive capabilities that enable early identification of bottlenecks and opportunities to optimize operational flows and allocations are well manifested in manufacturing through Predictive Maintenance, which has the capacity to reduce unplanned downtime by as much as 20-40% and reduce maintenance expenses by 10-30% (Mochalov, 2024). Better inventory management with predictive demand forecasts that reduce carrying cost and inventory stock outs, along with better quality management that reduces product defects, rework and warranty claim costs all impact operational efficiency and result in improved customer satisfaction. On the financial front, the improvements in operational performance impact revenue positively through improved customer segmentation and targeting, cross-selling and retention strategies, and have a large impact on the bottom line by reducing the cost of operations through areas such as fraud reduction, risk management and efficiency improvements (Wang & Li, 2026). With improved revenues and reduction of operating expenses as a consequence of the adoption and exploitation of BDA, profitability improves. However, there is, of course, a varying amount of return on investment depending on the value of the analytics initiatives, however, the correlation of the degree of use of advanced analytical capabilities to the business benefits is positive, as documented, among others, by a research on some 700 European companies, which shows that a higher use of analytical functions results in significantly greater benefits (Fabac, 2010).

6.3 Competitive Advantage and Innovation

BDPA enables competitive differentiation and innovation:

Customer Insight: Deeper understanding of customer needs and behaviors enables personalized experiences and relationship management.

Market Agility: Predictive capabilities enable faster responses to market changes, competitor moves, and emerging opportunities.

New Business Models: Analytics-driven insights have enabled entirely new business models, such as outcome-based services and dynamic pricing.

Innovation: Understanding of latent customer needs and market opportunities drives product and service innovation.

7. Future Directions and Open Issues

7.1 The Future of Data-Driven Decision Making

On the technological horizon, artificial intelligence (specifically deep learning and reinforcement learning)is achieving groundbreaking prediction accuracies, edge computing allows for near-real-time,low-latency analytics in highly distributed environments by conducting data processing and analysis closer to its source of collection (Benbya et al., 2020),while quantum computing promises a future with enhanced capabilities to solve previously intractable problems in optimization and simulations, although real-world adoption is still on the horizon;and autonomous systems are increasingly embedding predictions into self-decisioning frameworks to enable self-optimizing operations and prompt adaptive response (Bari & Ara, 2024). Beside the technology frontiers for speed and predictive performance, there exists a trade-off between model complexity and interpretability which has motivatedresearchon Explainable AI (XAI) and interpretable models, aimed at shedding light on opaque black-box behaviour to support human trust, accountability and responsible use cases;while visual analytics and fairness auditing help in visually exploring large-scale models and systematically checking their unbiased deployment. As growing concerns about individual privacy coupled with restrictive regulations shape the research and deployment of Big Data analytics, privacy-preserving analytics is gaining more momentum in recent years through several paradigms, such as differential privacy techniques which add statistically calculated noise to query results in order to maintain an individual's anonymity, while overall query accuracy remains high (using aggregate queries);federated learning approaches where different clients learn a shared prediction model from their local decentralized data without sharing it explicitly (making it ideal for data with strict sensitivity and privacy requirements such as health or financial data),;and more theoretically ambitious but practically limited privacy techniques such as homomorphic encryption and secure multi-party computation(Pathak et al., 2023), which promise computing on encrypted data as well as secure cross-organizational data collaborations without raw data being revealed (Gomaa, 2025).

7.2 Privacy-Preserving Analytics

Growing privacy concerns and regulatory requirements drive research into techniques enabling analytics while protecting sensitive information:

Differential Privacy adds mathematical noise to query results to protect individual privacy while preserving aggregate patterns.

Federated Learning enables model training across distributed data without sharing raw data, particularly valuable for healthcare and sensitive applications.

Homomorphic Encryption allows computation on encrypted data, though performance limitations currently restrict practical applications.

Secure Multi-Party Computation enables collaborative analysis across organizations without sharing proprietary data (Ali, 2022).

8. Conclusion

Our systematic review explored the state of big data predictive analytics (BDPA) for decision-making in complex business environments. By synthesizing evidence from 109 scientific articles, we gained a comprehensive understanding of its capabilities, applications, challenges, and outcomes. Our research contributes both to the academic knowledge and the practice of organizations undergoing the data-driven transformation. Big data phenomena are summarized as a "5V" model (volume, velocity, variety, veracity, value), describing the breadth and complexity of the decision problem, whereas analytics types are categorized in an "analytics continuum" of Descriptive, Diagnostic, Predictive and Prescriptive approaches towards the optimum level of analytical maturity. Secondly, a BDPA application domain taxonomy of seven key business areas indicates the extent to which prediction is used and to which extent those areas Benefit From BDPA applications (BM). Manufacturing is among the areas showing great Benefits Of Using BDPA, in forms of predictive maintenance and demand forecasting. Predictive modelling also plays a significant role in healthcare in areas of patient readmission prediction and in the finance industry, such as predicting defaults. A range of application diversity implies a generalizability of the technology across domains, yet highlights distinct applications characteristics in each area. Thirdly, the variety of predictive analytical techniques available including classical statistical methods, and Machine Learning (ML) and Deep Learning (DL) architectures. Given the underlying prediction mathematical formulation, the comparative analysis of these techniques enables organizations to select the ones that best fit their requirements, with a note that the emergence of more advanced methodologies has increased its capabilities at the expense of explainability and technical infrastructure. Fourth, implementation challenges of BDPA were explored. The evidence points to significant issues regarding data and analytics governance, including quality, availability, security and privacy. These challenges lead to decreased performance of prediction systems and raise concerns around ethical, legal and compliance aspects of decision-making. Other factors contributing to slow adoption are knowledge and skills gaps, employee's resistance and technical scalability. Fifth, compelling evidence regarding the business benefits and advantages that Organizations Get From Bdpa were identified across the research. The financial benefits are apparent, reflected in terms of improved operational and business performance, such as reducing costs and improving business profits, leading ultimately to long-term competitive advantages. For example, benefits of up to 40% in reducing planned downtime, improving service levels through a reduction of inventory and customer retention are reported. A comprehensive theoretical framework integrating the described findings (including the interactions between BDA capabilities, decision-making and business performance as mediated by various factors and moderated by business complexities) was proposed that serves both as guidance for future research and a framework for practical application by organizations undergoing this journey. A few important future directions need to be investigated to advance BDPA beyond its current limits. There is an ever-increasing demand for Explainable AI (XAI) systems to ensure human understanding, trust and fairness when relying on algorithmically-derived decisions. Another promising research area concerns privacy-preserving analytics such as federated learning, differential privacy and homomorphic encryption as organizations use more sensitive data in predictive models and are bound by stricter privacy laws. Further integration of prediction with autonomous systems, enabled by AI, to fully automate decision-making, although still facing significant hurdles in ensuring accountability and reliability, would lead to even more efficient and adaptable operations. Ultimately, further investigation of the human and organizational factors crucial to successful BDPA adoption will remain the most important theme for making the translation of predictive technologies to value and long-term business success a reality. In conclusion, BDPA is an integral approach offering Organizations Advantages Of Utilizing predictive modelling and decision support tools that enable them to manage complexity and make decisions based on anticipation and evidence. This, in turn, is a cornerstone for their transformation in adata-driven future with sustainable performance gains and competitive advantage, provided that organizations consistently focus on technology, data, skills, ethics and human engagement to unleash the full potential of data-driven intelligence.

REFERENCES

Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, *9*(1), 107.

Ali, H. (2022). Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making. *Int J Comput Appl Technol Res*, *11*(3), 88-104.

Bari, M. D., & Ara, A. (2024). The impact of machine learning on prescriptive analytics for optimized business decision-making. *Anjuman, The Impact Of Machine Learning On Prescriptive Analytics For Optimized Business Decision-Making (April 15, 2024)*.

Basu, D., & Das, U. (2025). The fair game: Auditing & debiasing AI algorithms over time. *Cambridge Forum on AI: Law and Governance*, *1*, e27.

Belani, G. (2025). Big data and predictive analytics: A systematic review of applications. *IEEE Computer Society*.

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS quarterly*, *44*(1), 1-18.

Bonthu, C. (2025). The role of data governance in strengthening ERP and MDM collaboration. *IJCESEN*.

Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Chaudhary, P. S., Khurana, M. R., & Ayalasomayajula, M. (2024). Real-world applications of data analytics, big data, and machine learning. In *Data Analytics and Machine Learning: Navigating the Big Data Landscape* (pp. 237-263). Singapore: Springer Nature Singapore.

Choudhury, S. J., & Pal, N. R. (2021). Classification of incomplete data integrating neural networks and evidential reasoning. *Neural Computing and Applications*.

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision support systems*, *55*(1), 359-363.

Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the Big Data Ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE.

Demirbaga, Ü., Aujla, G. S., Jindal, A., & Kalyon, O. (2024). Big Data Analytics. In *Big Data Analytics: Theory, Techniques, Platforms, and Applications* (pp. 31-42). Cham: Springer Nature Switzerland.

Demirbaga, Ü., Aujla, G. S., Jindal, A., & Kalyon, O. (2024). Big Data Analytics. In *Big Data Analytics: Theory, Techniques, Platforms, and Applications* (pp. 31-42). Cham: Springer Nature Switzerland.

Fabac, R. (2010). Complexity in organizations and environment-adaptive changes and adaptive decision-making. *Interdisciplinary Description of Complex Systems: INDECS*, *8*(1), 34-48.

Glang, F., Fabian, M. S., German, A., Khakzar, K. M., Mennecke, A., Liebert, A., ... & Zaiss, M. (2023). Linear projection-based chemical exchange saturation transfer parameter estimation. *NMR in Biomedicine*, *36*(6), e4697.

Gomaa, A. H. (2025). Enhancing Business Operations through Data-Driven Decision Making: A Comprehensive Review, Research Gaps, and Strategic Framework. *Journal of Decision Science and Optimization*, *1*(1), 84-96.

Hasanuddin, M. (2025). Enhancing Data Processing Efficiency through Machine Learning Algorithms: A Comprehensive Study. *Journal of Data Science and Informatics Engineering*, *1*(1), 1-9.

Hikmawati, S. (2021). Improving data quality and data governance through master data management. *IJITEE*, 5(3).

Islam, M. R., & Ikbal, M. Z. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review of Studies Across Retail, Finance, and Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62.

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). Big data and predictive analytics: A sytematic review of applications. *Artificial intelligence review*, *57*(7), 1.

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). Big data and predictive analytics: A systematic review of applications. *Artificial Intelligence Review*, *57*(7), 176. https://doi.org/10.1007/s10462-024-10811-5

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). Big data and predictive analytics: A sytematic review of applications. *Artificial intelligence review*, *57*(7), 1.

Leon, M. (2026). Lifecycle-based governance to build reliable ethical AI systems. *Systems Research and Behavioral Science, 43*(3), 1116–1132.

Liu, Y., & Wu, Y. (2007). Variable selection via a combination of the l 0 and l 1 penalties. *Journal of Computational and Graphical Statistics*, *16*(4), 782-798.

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2026). e-profits: a business-aligned evaluation metric for profit-sensitive customer churn prediction. *International Journal of Data Science and Analytics*, *22*(1), 75.

Masache, S., & Tanner, W. (2026, April 1). Marketing analytics and data privacy in 2026. Verndale.

Michael, I. (2025). Evaluation metrics and benchmarking for predictive accuracy.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. *British journal of management*, *30*(2), 272-298.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information systems and e-business management*, *16*(3), 547-578.

Mochalov, A. N. (2024). Digital profiling and the legal regime of derived personal data. *Kutafin Law Review, 11*(3), 496–505.

Moreira, M. W., Rodrigues, J. J., Korotaev, V., Al-Muhtadi, J., & Kumar, N. (2019). A comprehensive review on smart decision support systems for health care. *IEEE Systems Journal*, *13*(3), 3536-3545.

Noyes, J., Cook, M., & Masakowski, Y. (Eds.). (2012). *Decision making in complex environments*. Ashgate Publishing, Ltd..

Nwaimo, C. S., Oluoha, O. M., & Oyedokun, O. Y. E. W. A. L. E. (2019). Big data analytics: technologies, applications, and future prospects. *Iconic Research and Engineering Journals*, *2*(11), 411-419.

Pathak, S., Krishnaswamy, V., & Sharma, M. (2023). Big data analytics capabilities: a novel integrated fitness framework based on a tool-based content analysis. *Enterprise Information Systems*, *17*(1), 1939427.

Ponte, E. (2026). Reinforcement Learning: A Paradigm Shift in AI Training and its Competitive Implications. *Journal of Business & Technology Law*, *21*(1), 3.

Rane, N. L., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). Machine learning and deep learning for big data analytics: A review of methods and applications. *Partners Universal International Innovation Journal*, *2*(3), 172-197.

Sauter, V. L. (2014). *Decision support systems for business intelligence*. John Wiley & Sons.

Shafa, H. (2025). Artificial intelligence-driven business intelligence models for enhancing decision-making in us enterprises. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 771-800.

Sheng, T., Chen, Y., & Wang, H. (2025). Ethical challenges in the algorithmic era: A systematic rapid review of risk insights and governance pathways for nursing predictive analytics and early warning systems. *BMC Medical Ethics, 26*, 151.

Shin, D. H., & Choi, M. J. (2015). Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, *32*(2), 311-320.

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard business review*, *85*(11), 68.

Sodero, A., Jin, Y. H., & Barratt, M. (2019). Data and predictive analytics use for logistics and supply chain management. *International Journal of Physical Distribution & Logistics Management*, 49(7), 706-726.

Wang, Y., & Li, J. (2026). From ethical principles to executable governance: A policy-as-code framework for trustworthy AI in higher education. *Journal of Responsible Technology, 18*, 100061.