Data Intelligence and Informatics Practice

Research Article
Volume 1(1), 2026, 32−41
https://doi.org/10.53623/diip.v1i1.1267

Adoption and Sectoral Impact of Data Science Practices in India: A Cross-Sectional Case Study

Karthik Subramanian*, Prasanna S. Reddy

Sahyadri College of Engineering & Management, Sahyadri Campus, Adyar, Mangaluru - 575007, Karnataka, India

*Correspondence: karthiksubramanian988@gmail.com

SUBMITTED: 18 June 2026; REVISED: 26 July 2026; ACCEPTED: 31 July 2026

ABSTRACT: Data science has become a defining capability for organisations seeking to convert large volumes of structured and unstructured information into actionable decisions, yet the extent and pattern of its adoption across India's diverse industrial landscape remain incompletely documented. This study reports a cross-sectional survey of 420 organisational respondents drawn from seven sectors—information technology and IT-enabled services (IT/ITES), banking, financial services and insurance (BFSI), healthcare, government and public sector, agriculture and allied industries, retail/e-commerce, and education—across metro and non-metro locations in India. A structured, validated questionnaire was used to capture adoption levels, tools and techniques in use, perceived barriers, and organisational impact. Results showed that 55.0% of organisations had reached at least an intermediate level of data-science adoption, with marked sectoral disparity: IT/ITES and BFSI organisations reported the highest adoption (81.3% and 72.6% intermediate-to-advanced, respectively), while agriculture and allied industries and education lagged considerably (21.3% and 23.4%). Python, SQL and spreadsheet tools dominated the technology stack, while distributed big-data frameworks and natural-language-processing tools remained comparatively underused. Shortage of skilled manpower (74.3%), data-privacy concerns (65.2%) and poor data quality (52.6%) were the most frequently cited barriers. Adopting organisations reported the strongest perceived gains in decision-making (mean = 4.21/5) and operational efficiency (mean = 4.05/5). A five-year retrospective trend indicated accelerating adoption across all sectors between 2020 and 2025, with the gap between leading and lagging sectors widening rather than narrowing. These findings suggest that targeted workforce development, clearer data-governance frameworks and infrastructure investment in lagging sectors are needed to translate India's data-science momentum into more broadly distributed organisational value.

KEYWORDS: Data science; big data analytics; India; sectoral adoption; case study; digital transformation

1. Introduction

Data science is the interdisciplinary practice of using statistical reasoning, computational algorithms and domain knowledge to convert raw data into actionable insight that supports evidence-based decision-making [1]. Since the term began separating itself conceptually from classical statistics and database management, organisations worldwide have invested heavily in the people, platforms and governance structures needed to operationalise data-driven strategies [1]. India occupies a distinctive position in this global transformation: its rapidly expanding digital infrastructure, near-universal mobile connectivity and government-led digitalisation initiatives have made it one of the largest generators of structured and unstructured data in the world, while simultaneously exposing wide disparities in how different sectors translate this data into value.

Empirical evidence from Indian manufacturing and service-sector firms indicates that big-data adoption is shaped jointly by technological readiness, top-management support, competitive pressure and data-privacy concerns, and that adoption levels vary considerably even within the same industry [2]. At a conceptual level, big-data analytics continues to face data-related, process-related and management-related obstacles that cut across sectors and geographies, including issues of volume, veracity, integration and governance [3]. In the Indian health sector specifically, big-data analytics has been identified as holding substantial promise for improving diagnosis, treatment planning, hospital operations and public-health surveillance, yet practical uptake remains uneven across hospitals, insurers and pharmaceutical companies [4–6]. Parallel work on healthcare informatics has documented a rapidly growing but fragmented body of research linking big data to clinical and administrative outcomes, underscoring the need for sector-specific empirical assessment rather than generic technology narratives [7].

Agriculture, which still employs a substantial share of India's workforce, has likewise begun to draw on machine-learning techniques for tasks such as rainfed crop-yield forecasting, although the comparative performance of different algorithms and the readiness of smallholder ecosystems to absorb such tools remain open questions [8]. In banking and financial services, big-data analytics has been used for fraud detection, customer segmentation and risk management, with Indian commercial and public-sector banks reported to be catching up with international counterparts but still facing structural and skill-related constraints [9,10]; complementary evidence from the fintech sector suggests that digital financial services have meaningfully expanded financial inclusion, particularly for previously underbanked populations [11]. Government bodies have similarly experimented with data-driven citizen engagement, for example using transformer-based natural-language-processing models to automatically triage grievances submitted through social media to Indian Railways [12], while national-level analytics were used during the COVID-19 pandemic to model and predict outbreak trajectories across Indian states and union territories [13]. Beyond specific applications, country-level analysis suggests that artificial-intelligence and data-driven technologies are already exerting a measurable, if unevenly distributed, influence on India's broader economic growth and employment patterns [14].

Despite this expanding application base, India's emerging data-governance architecture, including evolving frameworks for personal and non-personal data protection, continues to create uncertainty for organisations seeking to scale analytics responsibly [15], and privacy-preservation techniques for big-data systems remain an active and unresolved area of research both globally and within the Indian context [16,17]. Taken together, the existing literature indicates that data-science adoption in India is sectorally uneven, methodologically fragmented and only partially understood at the level of practising organisations rather than isolated case examples. What remains comparatively scarce is a single empirical assessment that simultaneously profiles the extent of adoption, the tools in use, the obstacles encountered and the perceived organisational impact across multiple sectors using one consistent survey instrument. The present study addresses this gap through a cross-sectional, multi-sector survey of professionals and organisations operating in India, with the specific objectives of (i) describing the demographic and organisational profile of respondents engaged in data-science-related work, (ii) quantifying sector-wise adoption levels and the tools and techniques in use, (iii) identifying the principal barriers to wider adoption, and (iv) evaluating the perceived organisational impact of data-science practices and how this has evolved over a five-year period.

2. Materials and Methods

2.1. Study Design and Setting.

This study employed a cross-sectional, descriptive-analytical survey design to examine data-science adoption among organisations operating in India. Data were collected between January and June 2025 through a structured, self-administered online questionnaire. The study was not restricted to a single state; respondents were drawn from metropolitan centres (Delhi National Capital Region, Mumbai, Bengaluru, Hyderabad, Chennai, Pune and Kolkata) and from a purposive set of tier-2/non-metro cities, allowing comparison between metro and non-metro contexts.

2.2. Study Population and Sampling Technique.

The target population comprised professionals and organisational representatives—including data scientists/analysts, information-technology personnel, business/operations managers, academic researchers and government officials—engaged in or responsible for evaluating data-related technologies within their organisation. A multi-stage sampling strategy was used: organisations were first stratified by sector (IT/ITES, BFSI, healthcare, government/public sector, agriculture and allied industries, retail/e-commerce, and education), after which respondents within each stratum were recruited through professional networking platforms and industry-association mailing lists using simple random sampling. The minimum required sample size was estimated using Cochran's formula for an assumed population proportion of 0.5, a 95% confidence level and a 5% margin of error, yielding a minimum requirement of 385 respondents; to allow for incomplete responses, 460 invitations were issued, of which 420 fully completed and valid responses were retained for analysis (an effective response rate of 91.3%).

2.3. Data Collection Instrument.

Data were collected using a structured questionnaire developed after a review of the relevant literature [2,3,9] and refined through a pilot test with 30 respondents who were not included in the final sample. The instrument comprised four sections: (a) demographic and organisational characteristics; (b) current level and nature of data-science adoption, including specific tools, platforms and techniques in use; (c) barriers and challenges to adoption; and (d) perceived organisational impact, measured on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). Internal consistency of the perceived-impact scale was acceptable (Cronbach's $\alpha = 0.84$). The questionnaire was administered in English through a secure online survey platform.

2.4. Study Variables.

The independent/explanatory variables included sector of operation, organisation size, geographic region (metro versus non-metro) and respondent role. The primary outcome variables were (i) the categorical level of data-science adoption (not adopted, basic, intermediate, advanced), determined from a composite of self-reported usage-frequency and capability items, and (ii) the continuous perceived-impact score derived from the Likert-scale items in section (d) of the instrument.

2.5. Data Analysis Procedure.

Completed responses were coded and analysed using IBM SPSS Statistics (version 27.0). Descriptive statistics (frequencies and percentages) were used to summarise demographic, organisational and adoption-related variables. The association between sector and categorical adoption level was tested using the chi-square test of independence. Differences in mean perceived-impact scores across sectors were assessed using one-way analysis of variance (ANOVA) with Tukey's post-hoc comparison. A two-tailed p-value of less than 0.05 was considered statistically significant throughout. Trend data on adoption over the 2020–2025 period were obtained through a single retrospective recall item and are reported descriptively, in line with the limitations of self-reported historical data.

2.6. Ethical Considerations.

The study protocol was reviewed and approved by the Institutional Research Ethics Committee prior to data collection (approval reference: IREC/2025/DS-014). Participation was entirely voluntary, and electronic informed consent was obtained from all respondents before the survey was administered. No personally identifying information was collected, and all responses were anonymised at the point of entry to protect respondent confidentiality, consistent with prevailing data-protection norms [15].

3. Results and Discussion

3.1. Demographic and Organizational Profile of Respondents.

Table 1 summarises the demographic and organisational characteristics of the 420 respondents. The sample was predominantly male (61.0%), with the majority aged between 21 and 40 years (74.8%), reflecting a relatively young and economically active workforce. Most respondents held a master's degree or higher (60.0%), indicating a well-educated participant pool with substantial professional qualifications. In terms of industry representation, respondents were drawn primarily from the IT/ITES (22.9%) and banking, financial services, and insurance (BFSI) (20.0%) sectors, alongside participants from a range of other public and private industries. Geographically, respondents were more likely to be based in metropolitan cities (62.1%) than in non-metro or tier-2 locations (37.9%), consistent with the concentration of digital infrastructure and organisational innovation in India's major urban centres. Overall, the sample provides broad sectoral representation and captures respondents with diverse organisational backgrounds; however, the predominance of metropolitan and highly educated participants suggests that rural, small-scale, and informal-sector organisations may be somewhat under-represented.

Table 1. Demographic and Organizational Profile of Survey Respondents (N = 420).

Variable	Category	n (%)
Gender	Male	256 (61.0)
	Female	158 (37.6)
	Prefer not to say	6 (1.4)
Age Group (years)	21–30	168 (40.0)
	31–40	146 (34.8)
	41–50	79 (18.8)
	>50	27 (6.4)
Highest Education	Bachelor's degree	142 (33.8)
	Master's degree	211 (50.2)
	Doctoral degree	41 (9.8)
	Diploma/Other	26 (6.2)
Current Job Role	Data scientist/analyst	134 (31.9)
	IT/software engineer	98 (23.3)
	Business/operations manager	87 (20.7)
	Academic/researcher	54 (12.9)
	Government official	47 (11.2)
Organization Size (employees)	<50	76 (18.1)
	50–250	121 (28.8)
	251–1000	134 (31.9)
	>1000	89 (21.2)
Sector	IT/ITES	96 (22.9)
	BFSI	84 (20.0)
	Healthcare	61 (14.5)
	Government/Public Sector	58 (13.8)
	Agriculture & Allied	47 (11.2)
	Retail/E-commerce	44 (10.5)
	Education	30 (7.1)
Region	Metro city	261 (62.1)
	Non-metro/Tier-2 city	159 (37.9)
Total Respondents	—	**420 (100.0)**

3.2. Sector-wise Adoption of Data Science Practices.

Table 2 and Figure 1 present the distribution of self-reported data-science adoption levels by sector. Overall, more than half of the surveyed organisations (55.0%) had reached at least an intermediate level of adoption, while 18.1% reported no structured data-science practice at all. Adoption level was significantly associated with sector ($\chi^2(18) = 132.74$, $p < .001$). IT/ITES organisations reported the highest combined intermediate-to-advanced adoption (81.3%), consistent with the technological-readiness and competitive-pressure mechanisms identified in earlier work on Indian big-data adoption [2], whereas agriculture and allied industries lagged considerably (21.3% intermediate-to-advanced), reflecting the infrastructural and human-capital constraints widely reported for India's agricultural data ecosystem [8]. The relatively higher adoption observed in BFSI (72.6% intermediate-to-advanced) is consistent with earlier accounts of Indian banks adopting big-data analytics for fraud detection, customer segmentation and risk management [9,10], while the comparatively modest figures for government/public-sector bodies (31.0%) suggest that automated citizen-engagement tools, although piloted successfully in specific contexts such as railway-grievance handling [12], have not yet been scaled uniformly across departments.

Table 2. Sector-wise Distribution of Data Science Adoption Levels Among Surveyed Organizations (N = 420).

Sector	n	Not Adopted, n (%)	Basic, n (%)	Intermediate, n (%)	Advanced, n (%)
IT/ITES	96	4 (4.2)	14 (14.6)	31 (32.3)	47 (49.0)
BFSI	84	6 (7.1)	17 (20.2)	33 (39.3)	28 (33.3)
Healthcare	61	12 (19.7)	22 (36.1)	19 (31.1)	8 (13.1)
Government/Public Sector	58	19 (32.8)	21 (36.2)	13 (22.4)	5 (8.6)
Agriculture & Allied	47	21 (44.7)	16 (34.0)	8 (17.0)	2 (4.3)
Retail/E-commerce	44	3 (6.8)	11 (25.0)	18 (40.9)	12 (27.3)
Education	30	11 (36.7)	12 (40.0)	5 (16.7)	2 (6.7)
Overall	**420**	**76 (18.1)**	**113 (26.9)**	**127 (30.2)**	**104 (24.8)**

Note: Percentages are calculated within each sector. Data science adoption levels differed significantly across sectors ($\chi^2(18) = 132.74$, $p < 0.001$).

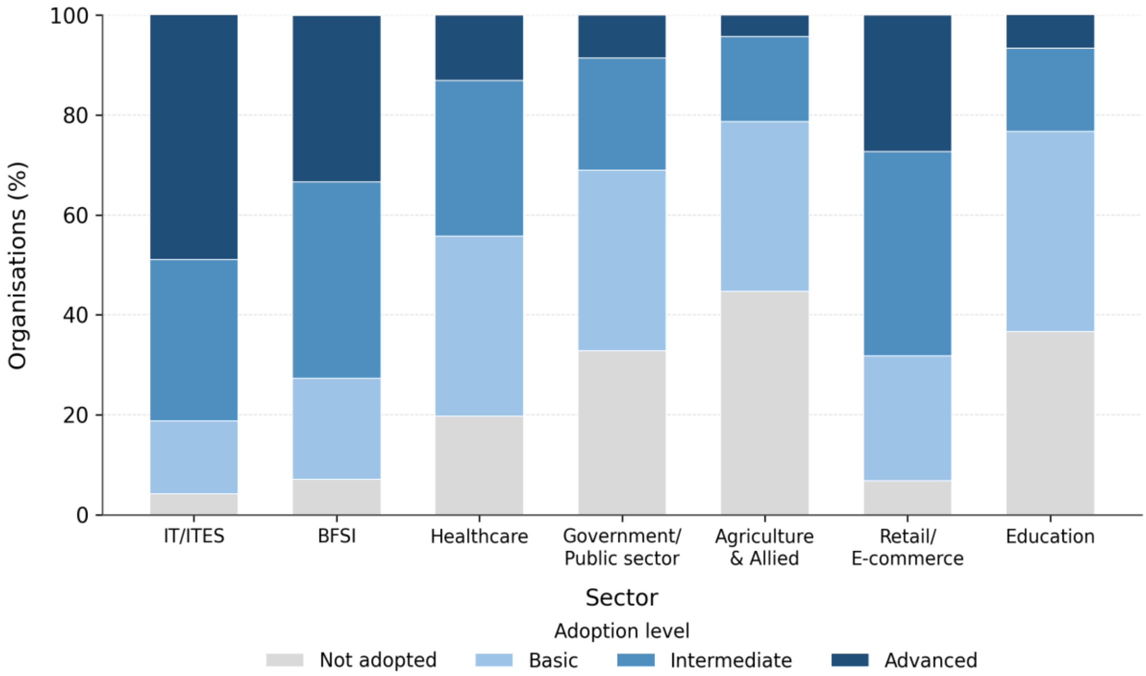

Figure 1. Sector-wise Distribution of Data Science Adoption Levels Among Surveyed Organisations in India (N = 420).

3.3. Tools, Platforms and Techniques Used in Data Science Practice.

Among the 344 organisations reporting at least a basic level of adoption, Python was the most widely used language (77.9%), followed closely by SQL/relational databases (70.1%) and Microsoft Excel (66.9%); the continued prominence of spreadsheet-based analysis alongside programming languages indicates that many organisations operate a hybrid analytics stack rather than a fully modernised one. Business-intelligence and visualisation tools such as Power BI and Tableau were used by 57.6% of adopters, and cloud platforms (AWS, Azure or Google Cloud) by 51.2%, reflecting the broader shift toward cloud-hosted analytics infrastructure. Distributed big-data frameworks such as Hadoop and Apache Spark (28.2%) and natural-language-processing tools (22.7%) were comparatively underused, suggesting that adoption in most surveyed organisations remains concentrated at the descriptive and diagnostic end of the analytics spectrum rather than at the predictive or prescriptive end, a pattern broadly consistent with persistent gaps between data-science rhetoric and applied analytical sophistication noted elsewhere [1,3].

Table 3. Data Science Tools, Platforms, and Techniques Used by Organizations (n = 344; multiple responses permitted).

Rank	Tool / Platform / Technique	Users, n	Usage (%)
1	Python	268	77.9
2	SQL and Relational Databases	241	70.1
3	Microsoft Excel	230	66.9
4	Power BI / Tableau	198	57.6
5	Cloud Platforms (AWS, Azure, GCP)	176	51.2
6	Machine Learning Frameworks (Scikit-learn, TensorFlow, PyTorch)	142	41.3
7	R Programming	134	39.0
8	Big Data Frameworks (Hadoop, Spark)	97	28.2
9	Natural Language Processing Tools	78	22.7

3.4. Barriers and Challenges to Data Science Adoption.

Table 4 summarises the barriers reported across the full sample. The most frequently cited obstacle was the shortage of skilled data-science manpower (74.3%), followed by data-privacy and security concerns (65.2%) and poor data quality or fragmented data sources (52.6%). The prominence of skill shortages corroborates earlier observations regarding the gap between industry demand and the available pool of professionals trained in the statistical and programming foundations underlying data science. The substantial proportion of respondents citing regulatory uncertainty linked to India's Digital Personal Data Protection framework (33.8%) echoes broader concerns documented in the legal and policy literature regarding the evolving and, in places, ambiguous nature of the country's personal and non-personal data-governance architecture [15], as well as the unresolved technical challenge of preserving privacy within large, heterogeneous datasets [16,17].

Table 4. Reported barriers to data science adoption among surveyed organizations (N = 420; multiple responses permitted).

Rank	Barrier	n	%
1	Shortage of skilled data science personnel	312	74.3
2	Data privacy and security concerns	274	65.2
3	High infrastructure and technology costs	238	56.7
4	Poor data quality and fragmented data sources	221	52.6
5	Lack of top management support or strategic vision	196	46.7
6	Inadequate digital infrastructure	187	44.5
7	Regulatory uncertainty regarding data protection	142	33.8
8	Resistance to organizational and cultural change	119	28.3

3.5. Perceived Organizational Impact and Five-Year Adoption Trend.

Among organisations that had adopted data-science practices, the impact indicator rated most positively was improved decision-making (mean = 4.21, SD = 0.68), followed by increased operational efficiency (mean = 4.05, SD = 0.74); cost reduction received the lowest, though still broadly positive, rating (mean = 3.55, SD = 0.85), as shown in Table 5. Perceived impact differed significantly by sector ($F(6,337) = 8.42$, $p < .001$), with BFSI and IT/ITES respondents reporting the highest composite impact scores, a pattern that aligns with their comparatively mature adoption levels reported in Section 3.2 and with documented gains in fraud detection and operational efficiency in financial big-data applications [9–11]. The five-year trend depicted in Figure 2 shows a clear and accelerating upward trajectory across all sectors between 2020 and 2025, with IT/ITES rising from 38% to 86% and government/public-sector adoption rising more modestly from 9% to 44%; the steeper growth of the BFSI and IT/ITES trajectories relative to agriculture and government indicates that the gap between leading and lagging sectors has widened rather than narrowed over the period. The acceleration evident from 2023 onward broadly coincides with intensified national digitalisation efforts and growing investment in artificial-intelligence-related technologies reported at the macro-economic level [14], consistent with country-level evidence that such technologies are beginning to exert a measurable, though sector-uneven, influence on Indian economic activity.

Table 5. Perceived Organizational Impacts of Data Science Adoption Among Adopting Organizations (n = 344).

Rank	Impact Indicator	Mean	SD
1	Improved decision-making	4.21	0.68
2	Increased operational efficiency	4.05	0.74
3	Enhanced customer satisfaction/personalization	3.87	0.81
4	Better fraud and risk detection	3.74	0.93
5	Strengthened innovation capability	3.69	0.79
6	Revenue and profitability growth	3.62	0.88
7	Cost reduction	3.55	0.85

Note: Responses were measured on a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree

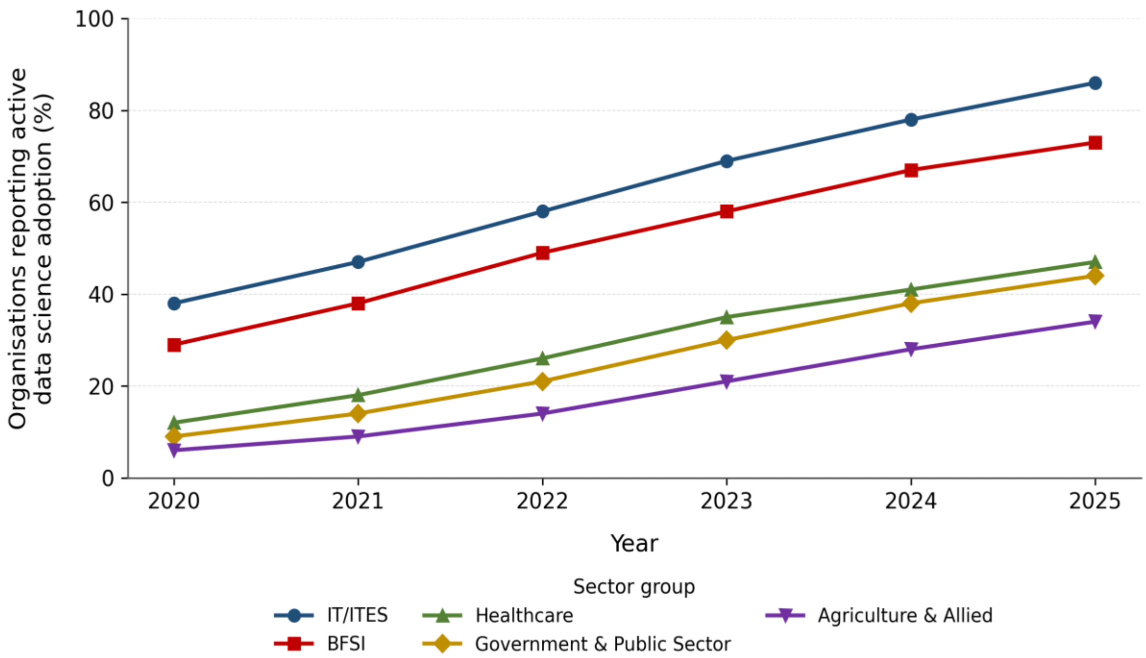

Figure 2. Five-year Trend in Data Science Adoption Across Major Sector Groups in India, 2020–2025, Based On Retrospective Self-Report (n = 344 adopting organisations).

4. Conclusion

This cross-sectional case study of 420 organisational respondents indicates that data-science adoption in India has progressed substantially over the past five years but remains markedly uneven across sectors, with IT/ITES and BFSI organisations operating at considerably higher levels of maturity than agriculture, education and government bodies. Skill shortages, data-privacy concerns and infrastructural constraints continue to limit the pace and depth of adoption, even as organisations that have adopted data-science practices report tangible improvements in decision-making, efficiency and risk management. Bridging the sectoral gap will likely require coordinated investment in workforce training, clearer data-governance frameworks, and infrastructure support tailored to lagging sectors such as agriculture and government administration. Because the study relies on self-reported, cross-sectional survey data from a sample recruited through professional networks, the findings should be interpreted as indicative of broad national patterns rather than as definitive sector-level estimates; future longitudinal and sector-specific research is recommended to validate and extend these results.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

Both authors contributed equally to the conceptualisation, methodology, data collection, formal analysis, and writing (original draft and review and editing) of this manuscript. Both authors have read and approved the final version of the manuscript.

Data Availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Provost, F.; Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508.

[2] Gangwar, H. (2018). Understanding the determinants of big data adoption in India: An analysis of the manufacturing and services sectors. *Information Resources Management Journal*, *31*(4), 1–22. https://doi.org/10.4018/IRMJ.2018100101.

[3] Sivarajah, U.; Kamal, M.M.; Irani, Z.; Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001.

[4] Duggal, R.; Khatri, S.K.; Shukla, B. (2016). Opportunities and challenges of using big data analytics in Indian healthcare system. *Indian Journal of Public Health Research & Development*, *7*(4), 238. https://doi.org/10.5958/0976-5506.2016.00226.6.

[5] Raghupathi, W.; Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *2*(1), 3. https://doi.org/10.1186/2047-2501-2-3.

[6] Kavitha, R.; Kannan, E.; Kotteswaran, S. (2016). Implementation of cloud based electronic health record (EHR) for Indian healthcare needs. *Indian Journal of Science and Technology*, *9*(3), 1–5. https://doi.org/10.17485/ijst/2016/v9i3/86391.

[7] Gu, D.; Li, J.; Li, X.; Liang, C. (2017). Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International Journal of Medical Informatics*, *98*, 22–32. https://doi.org/10.1016/j.ijmedinf.2016.11.006.

[8] Karbassi Yazdi, A.; Durán, C.; Derpich, I.; González, G.V. (2026). Forecasting crop yields in rainfed India: A comparative assessment of machine learning baselines and implications for precision agribusiness. *Agriculture*, *16*(1), 65. https://doi.org/10.3390/agriculture16010065.

[9] Srivastava, U.; Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. *Procedia Computer Science*, *50*, 643–652. https://doi.org/10.1016/j.procs.2015.04.098.

[10] Nobanee, H.; Dilshad, M.N.; Al Dhanhani, M.; Al Neyadi, M.; Al Qubaisi, S.; Al Shamsi, S. (2021). Big data applications the banking sector: A bibliometric analysis approach. *Sage Open*, *11*(4). https://doi.org/10.1177/21582440211067234.

[11] Asif, M.; Khan, M.N.; Tiwari, S.; Wani, S.K.; Alam, F. (2023). The impact of fintech and digital financial services on financial inclusion in India. *Journal of Risk and Financial Management*, *16*(2), 122. https://doi.org/10.3390/jrfm16020122.

[12] Agarwal, S.; Kumar, A.; Ganguly, R. (2024). Investigating transformer-based models for automated e-governance in Indian Railway using Twitter. *Multimedia Tools and Applications*, *83*(2), 4551–4577. https://doi.org/10.1007/s11042-023-15331-y.

[13] Gupta, M.; Jain, R.; Arora, S.; Gupta, A.; Awan, M.J.; Chaudhary, G.; Nobanee, H. (2021). AI-enabled COVID-19 outbreak analysis and prediction: Indian states vs. union territories. *Computers, Materials & Continua*, *67*(1), 933–950. https://doi.org/10.32604/cmc.2021.014221.

[14] Panigrahi, A.K. (2024). Impact of artificial intelligence on Indian economy. *Journal of Management Research and Analysis*, *11*(1), 33–40. https://doi.org/10.18231/j.jmra.2024.007.

[15] Mishra, N.; Agrawal, B. (2026). The emerging framework for non-personal data protection in India: Perils, promises, and lessons for the developing world. *International Journal of Law and Information Technology*, *34*, eaaf025. https://doi.org/10.1093/ijlit/eaaf025.

[16] Rafiq, F.; Awan, M.J.; Yasin, A.; Nobanee, H.; Zain, A.M.; Bahaj, S.A. (2022). Privacy prevention of big data applications: A systematic literature review. *Sage Open*, *12*(2). https://doi.org/10.1177/21582440221096445.

[17] Ramakrishnan, R.; Sujithra, R.; Niranjalin, A.J. (2024). Privacy challenges and solutions in big data analytics: A comprehensive review. *International Journal for Research in Applied Science & Engineering Technology*. https://doi.org/10.22214/ijraset.2024.62103.