Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 1, 2025 (New Series), pp 225–247
https://www.ssca.org.in/journal

A Comprehensive Review of Data Science, Artificial Intelligence, and Big Data Analytics in Indian Official Statistics

Prasily P. and Manoharan M.
Department of Statistics, University of Calicut, Kerala, 673635

Received: 30 May 2023; Revised: 01 May 2024; Accepted: 09 June 2024

Abstract

This article examines how Data Science (DS), Artificial Intelligence (AI), and Big Data Analytics (BDA) are used in the Indian digital government to produce official statistics. Official statistics play a crucial role in shaping policy decisions, and scientific advances have made it possible to extract insights and patterns from vast amounts of data. The article examines the current state of official statistics in India and explores how the digital government is applying DS, AI, and BDA to upgrade statistical analysis. The article also discusses the challenges of executing these advanced technologies, including data quality and privacy concerns. Furthermore, the review highlights some recent developments and schemes for the benefit of DS, AI, and BDA in Indian official statistics, including the application of Machine Learning (ML) and predictive modeling. The article concludes with recommendations for future research and policy in this area, highlighting the need for equality between technological innovation and ethical considerations to ensure the precise and responsible use of official statistics in the Indian digital government.

Key words: Official statistics; Data Science; Artificial Intelligence; Big Data Analytics.

AMS Subject Classifications: 62A01, 97K80.

1. Introduction

Indian policy started to take shape from the days of colonial rule, with official statistics playing a key role. The British government started to collect and publish data on the Indian economy during the mid-nineteenth century. Data was collected on agriculture, demography, and trade from official records. The government continued to formulate and monitor economic and social policies based on official statistics after independence. The Central Statistical Office (CSO), established in 1951, became the nodal agency for collecting, compiling, and disseminating official statistics. (`https://unstats.un.org/unsd/wsd/docs/India_wsd_history.pdf`). Rao (2010) argues that India's rapid economic growth and ambitious development agenda make official statistics even more important.

Corresponding Author: Prasily P.
Email: prasilyprasi29@gmail.com

1.1. Background and significance of official statistics in India

India, which has a population of 1.42 billion, heavily depends on official statistics for policy-making decisions.(https://worldpopulationreview.com/countries, https://www.statista.com/statistics/263766/total-population-of-india/). Developing economic policies and programs in India is driven by official statistics. The Indian constitution mandates that the central government collect and publish official statistics associated with the country's economy, society, and population.

Official statistics are significant to Indian government officials and policymakers, and numerous steps have done to grow their availability and accuracy. The National Statistical System (NSS) gathers information on several socioeconomic factors via sample surveys. The NSS contains various organizations such as the CSO, the National Sample Survey Office (NSSO), and the Registrar General and Census Commissioner of India (RGCCI) (https://mospi.gov.in/142-present-indian-statistical-system-organisation).

The government uses official statistics for some functions, including planning, analyzing, and monitoring policies and programs. For instance, official statistics used to assess the progress of the Sustainable Development Goals (SDGs) and various flagship schemes of the government such as Sashakt Bharat - Sabal Bharat (Empowered and Resilient India), Swachh Bharat - Swasth Bharat (Clean and Healthy India), Samagra Bharat - Saksham Bharat (Inclusive and Entrepreneurial India), Satat Bharat – Sanatan Bharat (Sustainable India) and Sampanna Bharat- Samriddh Bharat (Prosperous and Vibrant India) (https://sustainabledevelopment.un.org/content/documents/26162Main_Messages_India.pdf). Non-governmental organizations, academics, and researchers are also using official statistics to study the economic facets.

The National Policy on Official Statistics (NPOS) (https://mospi.gov.in/sites/default/files/announcements/draft_policy_17may18.pdf) outlines the draft policy by the Government of India's Ministry of Statistics and Programme Implementation (MOSPI). It covers fundamental principles of official statistics, objectives, government policy initiatives, and mechanisms for regulating core statistics. This policy underlines the professional independence, confidentiality of data, and maintaining statistical standards to provide relevant and accurate empirical data to inform economic and social policies. It further addresses the decentralization of the statistical system in India and the involvement of various government bodies, including the MOSPI, Directorates of Economics & Statistics, and the National Statistical Commission. It also prioritizes ensuring quality, promoting data sharing, developing capacity, and cooperating internationally in official statistics. The revised draft (https://mospi.gov.in/sites/default/files/announcements/Draft_National_Policy_on_Official_Statistics.pdf) brings into perspective the transformative power of data and statistics in achieving sustainable development and inclusive growth. The initiatives embraced to reform and empower the institutional framework of the official statistical system in India are discussed, such as involving better coordination, international data standards adoption, SDG monitoring via the National Indicator Framework, and the Collection of Statistics Act amendment. It also indicates India's election to the United Nations Statistical Commission (UNSC) for 2024-2027 highlights its responsibility for maintaining global statistical efficiency and integrity.

Official data are significant, but gathering and analyzing them in India is fraught with

difficulties. The absence of timely and reliable data, especially in socio-economic indicators such as poverty and employment, represents a significant issue. The inconsistent definition and assessment of numerous indicators throughout the nation's states and regions is another problem. India's development planning and policymaking process both heavily rely on official statistics. The Indian government has taken measures to improve official statistics, but some issues still need to be resolved for reliable and timely data dissemination.

1.2. Overview of the role of digital government in official statistics

In recent years, the importance of using digital technologies and data analytics in official statistics has increased, and India is no exception. Digital government initiatives can enhance official statistics through better collection, analysis, and dissemination, leading to higher-quality, more timely, and comprehensive data.

The government of India has launched several initiatives to improve the availability and accessibility of official statistics using digital technologies. For example, the MOSPI launched the National Data Sharing and Accessibility Policy (NDSAP) (`https://dst.gov.in/national-data-sharing-and-accessibility-policy-0`) and the National Data and Analytics Platform (NDAP) (`https://ndap.niti.gov.in/`) to facilitate data sharing between different government departments and to improve the availability of official statistics.

NDSAPs main objective is to promote data sharing and reuse by defining standards for sharing and guidelines for data management. National governments have made their data available in public open formats for researchers, policymakers, and the general public.

NDAP is a web-based platform that opens up official statistics and datasets collected by different departments in the government. The platform aims to promote data-driven decision-making by making it easier for users to access and analyze official statistics. The most notable definiteness omitted from the platform is the appliance of data visualization and analysis.

Additionally, the Indian government has launched several other initiatives to improve the collection and analysis of official statistics through digital government. Mobile apps and cloud computing are examples of digital technologies used to improve data collection and analysis. The government is also exploring the application of data analytics and ML to automate data analysis and improve the accuracy and timeliness of official statistics.

Digital technology can transform the process of collecting, processing, and publishing official data, making it more accurate and accessible for policymakers. Utilizing digital technologies requires carefully evaluating data quality, privacy, and security concerns.

A particular focus will be placed on data science, AI, and BDA in India's official statistics as part of this review. The objectives of this review are to:

- Explore the historical and contemporary significance of official statistics in India.
- Identify the current challenges facing official statistics in India.
- Evaluate the potential of digital government in addressing these challenges.

- Analyze the implications of digital government for the reliability and quality of official statistics in India.
- Provide recommendations for future research and policy development in this area.

2. Overview of official statistics in India

Government agencies in India collect official statistics to inform policy and decision-making. The National Statistics Office (NSO) is responsible for providing official statistics covering population, economy, social welfare, natural resources, environment, and management. The data security, quality, privacy, accessibility, and coverage of data remain challenges despite progress. India's government is using digital technologies like NDAP and NDSAP to improve official statistics.

2.1. Brief history of official statistics in India

Official statistics in India have a long and intriguing history dating back to the colonial era. The CSO, founded in 1861 by the British colonial government, was the country of India's first official statistics office. The main objective of this agency was to provide statistical information for the British government's economic and administrative policies in India (Rao, 2010; Ghosh *et al.*, 1999)

After India gained independence in 1947, the Central Statistical Organization was renamed the Central Statistical Office and became part of the Ministry of Planning. The CSO's role has evolved to include gathering, compiling, and disseminating official statistics.

According to Sarma (1958), the government of India has taken pivotal measures to streamline the statistical system as part of its development of a system for official statistics. It supports economic aspects in planning implemented during the First Five Year Plan (1951-1956). In the consecutive five-year plan (1956-1961), National Sample Survey Offices (NSSOs) introduced the system to conduct surveys on different topics.

In 2005, the discussion of setting up the National Statistical Commission (NSC) was to provide guidance and direction to the statistical system in India (`https://mospi.gov.in/national-statistical-commission-0`). NSC was formed in 2006 to provide quality and integrity in statistics to the society. Rao (2013) discusses the NSC and its functions. It outlines the historical background of statistical data collection in the country and the role of the NSC in coordinating statistical activities. NSC is responsible for identifying core national statistics, formulating national policies related to the statistical system, and improving public confidence in official statistics. It has a pivotal role in shaping the official statistical system and meeting the statistical requirements of the nation.

The Indian government has prioritized the use of data analytics and digital technology in official statistics during the last few years. The government has started some programs, such as the NDSAP and NDAP, which were covered in the preceding section, to increase the availability and accessibility of official data. In general, the statistical system of India has gradually developed with an increasing focus on data for planning and policymaking. The government's focus on digital technology and data analytics expects change in future official statistics.

2.2. Current status and challenges of official statistics in India

The official statistics system in India has come a long way since its establishment. Despite efforts to improve the statistical system, several challenges remain.

The lack of readily available and high-quality data is one of the significant challenges facing official statistics in India. While India has made main progress in data collection and dissemination, there are still gaps in the data availability, particularly for specific social and economic indicators. It can also be problematic to determine the quality of the data collected; some data sources are not updated regularly or do not accurately capture the whole scope of the indicator being measured.

Another challenge is the better coordination and harmonization between different agencies involved in data collection and dissemination. Multiple agencies collect and publish data in India, and there is a lack of coordination between them, which can lead to inconsistencies and discrepancies in the data reported.

The dissemination of official statistics is also a challenge in India. Despite recent advances, official statistics are not always easily accessible to the general public, limiting their use and impact. Moreover, the statistical system must be more transparent to build trust in its accuracy and reliability.

The COVID-19 pandemic has highlighted some of these challenges, with the need for timely and accurate data becoming even more critical. The authors Hantrais *et al.* (2021) in the period of pandemic has highlighted the need for better investment in the statistical system to ensure its resilience and effectiveness in times of crisis.

To address these challenges, the Indian government has taken several steps to improve the statistical system, initiatives like the NDSAP and the NDAP aimed at increasing the availability and accessibility of official statistics. The government has also established a new NSC to provide guidance and direction to the statistical system and enhance the quality of official statistics.

Overall, while there have been significant improvements in the official statistics system in India, there is still a long way to go to ensure that official statistics are of high quality, accessible to all, and able to inform policy and decision-making effectively. For more details, refer to (`https://www.thehinducentre.com/publications/policy-watch/credible-data-for-public-good-constraints-challenges-and-the-way-ahead/article65971093.ece`).

2.3. Role of digital government in addressing these challenges

Official statistics in India face many challenges that the digital government can help address. The use of digital technologies and platforms can help improve data collection, quality, and dissemination, as well as increase transparency and accountability in the statistical system (Rana *et al.*, 2020; Chatterjee, 2020; Vijai, 2019).

One way in which digital government can improve official statistics is through the use of technology in data collection. Surveys conducted on mobile devices and online questionnaires can help increase the efficiency and accuracy of data collection, specifically in

hard-to-reach places. To reduce data entry errors and improve data collection efficiency, the Indian government has launched the NSSO mobile application. The use of digital technology can also help in real-time data collection, thereby ensuring the availability of timely data.

Another way in which digital government can improve official statistics is through the use of advanced analytics techniques such as BDA and ML. Using these techniques can uncover data insights missed by traditional statistics. India's NDAP integrates and analyzes administrative and survey data using BDA and AI.

Digital platforms, online portals, dashboards, and mobile applications can improve the accessibility and user-friendliness of official statistics. India's NDSAP promotes open data sharing and easy access to official statistics.

Furthermore, the digital government can play a role in increasing transparency and accountability in the statistical system. The use of digital platforms can help in the monitoring and reporting of statistical data and make it easier for stakeholders to identify any issues. Integrity and authenticity of official statistics can be ensured through blockchain, building trust.

To sum up, the digital government can improve official statistics in India. By utilizing digital technologies and platforms, the statistical system can become more efficient, accurate, and accessible, leading to better quality and impact of official statistics on policy and decision-making.

3. Digital government and official statistics in India

In India, the concept of digital government aims to use technology to enhance the effectiveness and efficiency of public services, including gathering and distributing official statistics. A digital government may transform official statistics collection, processing, and analysis with the help of big data and AI. Initiatives such as the NDAP and the NDSAP are examples of how the government of India is using digital technologies to improve the quality and accessibility of official statistics. Data security, privacy concerns, and the need to maintain data quality remain crucial factors for digital governance in India's official statistics (Alvarenga *et al.*, 2020; Tripathi and Dungarwal, 2020).

3.1. Role of data science in official statistics in India

Data science is playing an increasingly important role in official statistics in India. A large and complex dataset is analyzed using statistical and computational methods, including machine learning and predictive modeling. In India, the NSO is using data science to improve the accuracy and timeliness of official statistics. For more information, see Ashofteh and Bravo (2021).

NSO is developing a BDA platform to analyze various datasets, including agriculture (Sinha and Dhanalakshmi, 2022; Guntukula, 2020), health (Subrahmanya *et al.* (2022)), energy, and environment, etc. The platform aims to use advanced data science techniques to analyze large datasets to provide insights and inform policy decisions.

The use of data science in official statistics in India has the potential to enhance the

accuracy and efficiency of data collection and analysis. Data quality and privacy remain challenges, including the need for skilled professionals. The NSO has established guidelines and protocols for data privacy and confidentiality while also providing training programs to develop skills in data science.

3.2. Use of AI in official statistics in India

AI is also playing an increasingly important role in official statistics in India. AI involves the process and analysis of the data using computational models and algorithms, and it may automate several steps in data collecting and analysis (Chatterjee *et al.*, 2022; Vijai and Wisetsri, 2021).

In India, the NSO has been researching the use of AI in official statistics to upgrade the efficiency and accuracy of data collection and processing. For example, using ML algorithms to analyze satellite photos; for instance, may be used to estimate agricultural yields using AI. Ilyas *et al.* (2023).

The use of AI in official statistics in India has the potential to improve the accuracy and efficiency of data collection and analysis. The development and implementation of AI solutions still face many challenges, including data quality and privacy concerns Sharma *et al.* (2022). According to Ashofteh and Bravo (2021), the NSO has started training programs to advance expertise in AI and ML and has set rules and processes to guarantee data quality and privacy.

The National Institution for Transforming India (NITI) Ayog provides a report `https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf` focusing on the accelerated adoption of AI technology in India. AI plays a significant role in the national strategy, emphasizing the diversity, digital divide, scale, and lack of awareness in the Nation as factors that can amplify the risks associated with AI. In February 2021, an approach paper titled "Principles of Responsible AI" is scheduled, drawing on consultations and the Indian Constitution. AI can improve healthcare, agriculture, education, and entertainment, especially during the COVID-19 epidemic. It highlights how essential it is to use technology sensibly, reflecting the Prime Minister's remarks at the Davos Summit. It also discusses the necessity for a multidisciplinary approach to solve issues and foster confidence in AI systems, as well as the operationalization of responsible AI principles and the roles of the public, corporate, and research sectors.

3.3. Use of BDA in official statistics in India

The use of BDA in official statistics has become increasingly important in India Dubey *et al.* (2019a). BDA is a method for studying huge, complicated information to uncover insights and patterns that might help policymaking. In India, the NSO has been exploring the application of BDA to improve the accuracy and timeliness of data collection and analysis. The NSO has initiated a project to develop a comprehensive BDA platform for the research of various datasets, including those related to agriculture (Tantalaki *et al.*, 2019; Misra *et al.*, 2020), health (Chinnaswamy *et al.*, 2019; Li *et al.*, 2021), and the environment Dubey *et al.* (2019b, 2020). Nonetheless, issues like the lack of qualified personnel and the necessity to guarantee the privacy and quality of data continue to exist.

Despite facing challenges, BDA in official statistics has displayed promising outcomes in India. The BDA program has been used in India to evaluate and monitor various government programs, such as the Pradhan Mantri Jan Dhan Yojana (PMJDY) financial inclusion program (`https://pib.gov.in/Pressreleaseshare.aspx?PRID=1649091`) and the PMFBY crop insurance program (`https://transformingindia.mygov.in/scheme/pradhan-mantri-fasal-bima-yojana/`). Using BDA, the government accurately identified scheme beneficiaries and improved resource targeting. BDA is used to assess the effectiveness of environmental policies and programs, such as the Swachh Bharat Abhiyan (Clean India Mission) campaign (`https://www.pmindia.gov.in/en/major_initiatives/swachh-bharat-abhiyan/`). Consequently, BDA can facilitate data-driven policy decisions by the government by transforming Indian official statistics.

The goal of good governance in democratic countries is to ensure the provision of public services through effective participation to ensure accountability, responsiveness, and transparency. Meeting SDGs is one way of accomplishing this. Converged governance efforts at the grassroots level are required to achieve sustained development, which generates continuous baseline data. The amalgam of structured and unstructured data through BDA and emerging information and communication technologies (ICTs) can revolutionize governance processes and support data-backed decision-making Malhotra *et al.* (2018).

In addition, BDA can also help identify trends and patterns in official statistics that might otherwise go unnoticed. Analyzing social, economic, and environmental issues can be improved by combining data from social media, geospatial, and survey sources. BDA can be used for disease tracking and predicting crop yields based on weather patterns. As a result, the government can take proactive measures to address and prevent potential problems. For example, Mamatha *et al.* (2023) used BDA to track the spread of diseases, while Jaber *et al.* (2022) predicted crop yields using BDA. Policymakers can also use BDA insights to monitor and evaluate the effectiveness of government policies and programs, providing valuable data-driven feedback.

3.4. Exploring integrated data systems (IDS)

IDS are indispensable tools in National Statistics, ensuring data quality and efficient data collection for digital surveys. These systems aim to maximize information accessibility while minimizing user effort. The integration of data from various sources is a key aspect of the transformation of national statistics systems in the digital age Gokhberg *et al.* (2020). This integration should be supported by well-defined data governance frameworks Križman and Tissot (2022) and should consider the specific features of the digital economy Kasianova *et al.* (2021).

The integration process can be challenging, as emphasized by Sakshaug and Steorts (2023). Their discussion on merging surveys and administrative data underscores the complexities involved, including ethical considerations and computational burdens. Obtaining consent for data linkage and improving accuracy through computational techniques emerge as critical focal points.

Another innovative approach, as explored by Haim *et al.* (2023) involves a user-centered paradigm for data collection, where participants contribute digital trace data for academic research. This novel method combines survey data with donated data to unlock

deeper insights. Challenges such as methodological and ethical considerations are addressed alongside software solutions aimed at enhancing usability and reducing drop-outs.

The necessity for data and statistics in monitoring SDGs is addressed by Abbas *et al.* (2023). The authors highlight challenges in data dissemination and suggest AI as a potential solution. Furthermore, it proposes capacity development projects and a comprehensive indicator utilizing AI for processing data and producing official statistics.

The digital era brings challenges and opportunities, as discussed by Hassani and MacFeely (2023) Their comprehensive framework for digital data governance emphasizes the evolving landscape due to emerging technologies, underlining the importance of ethics and trustworthiness. Vavilova and Ketova (2023) developed an analytical system for regional socio-economic processes using official data and dynamic models to forecast and examine time, territory, and age-related indicators.

Daraio *et al.* (2022) proposes a completeness-aware integration approach to enhance data quality. Gootzen *et al.* (2023) introduces a quality framework for combining survey, administrative, and big data, showcasing its application in case studies involving mobility and virus detection data.

These studies elucidate the significance of IDS in shaping the future of National Statistics, offering valuable insights and innovative solutions to meet the challenges of the digital age.

3.5. Benefits and challenges of digital government in official statistics

Digital government has brought about many benefits in official statistics. Data science, AI, and BDA extract insights and patterns from vast data. As a result, statistical analysis is now more accurate and effective, which helps policymakers make better choices. But, there are also some difficulties with the help of the digital government in official statistics.

One of the benefits of digital government in official statistics is improved data quality. The utilization of digital systems guarantees the standardization and consistency of data collection, processing, and storage. As a result, data collection and analysis are more accurate, and statistical outputs are more reliable. Furthermore, digital systems enable real-time data collection, allowing for the latest statistical information essential in today's rapidly changing world.

Another benefit of digital government in official statistics is increased efficiency. Utilizing cutting-edge tools like ML and BDA speeds up data processing and analysis, requiring less time and effort than traditional statistical methods. Consequently, this enables policymakers to make more timely and informed decisions, leading to improved governance.

However, ensuring data privacy and security is one of the biggest challenges in adopting digital government in official statistics. Digital data storage requires data protection regulations and privacy laws to prevent breaches and misuse.

In today's society, a digital divide is growing between those with and without access to technology. India's digital government use in official statistics may leave behind marginal-

ized communities due to the significant digital divide, leading to unequal representation in statistical outputs. The need to overcome the digital divide and ensure that all communities have access to technology is evident from this. It is imperative to close the digital divide and ensure all communities have access to digital technology in light of these findings.

In conclusion, the benefits of digital government in official statistics are numerous, including improved data quality and increased efficiency. However, the approval of digital government also comes with several challenges, including data privacy and security and the digital divide. Balancing advanced technologies with ethical considerations is crucial for accessible digital government benefits in official statistics.

4. Case studies

4.1. Case study 1: Digital India and official statistics

Through the Digital India program, India aims to become a knowledge economy and a society empowered by digital technology. Improvements in official statistics will be a key focus of this initiative since they shape decisions regarding policy. This case study examines the implementation of digital India in official statistics and identifies its benefits and challenges.

4.1.1. Literature review

A knowledge-based economy is the objective of the Digital India initiative. Enhancing digital literacy and infrastructure and promoting digital services are essential to the program's success. Furthermore, digital technology will be essential to raising the standard of official data.

The use of technology in official statistics has been a topic of interest in the literature for several years. The study by Saxena (2018) explores the impact of demographic variables on the perception of corruption in e-government services in India. Hierarchical regression analysis shows that only gender influences the perception of corruption, with men perceiving a decrease and women perceiving an increase post-launch of the Digital India initiative. The study fills a gap in the literature by highlighting the importance of considering demographic variables in understanding citizens' perceptions of corruption in developing countries. Its small sample size and narrow focus on demographic variables limit the study.

The article by Rao (2019) examines the processes of identity creation in digital India through the use of Aadhaar. It challenges the distinction between identification and identity and shows how Aadhaar procedures create or deny conditions for belonging. It involves stitching together a digital signature, documentary proof of identity, and personal recognition to become a rights-bearing individual. Aadhaar adds a new layer of procedures on top of older methods of recognition, insisting on unique individual recognition while also recognizing a specific status.

Aadhaar is India's biometric program, which captures iris scans, fingerprints, facial photos, and demographic data from over 90% of the population. Nair (2021) argues that Aadhaar prompts a re-evaluation and contestation of individualism in postcolonial India because it dataficates the body. Additionally, it suggests that the program facilitates belonging

in the emerging technocratic imagination of a digital India.

The authors Gautam *et al.* (2022) conducted a study to examine the impact of financial technology on digital literacy in India, using the poverty score as a moderating variable. They found that Kisan Credit Cards (KCCs) had a positive association with literacy rate, while ATMs had a negative one. However, both KCCs and ATMs had a beneficial effect on literacy when interacting with poverty scores. The study's findings have implications for policymakers to understand the situation at the ground level while forming new policies for society's betterment. The authors suggest that ordinary people should take advantage of financial technology and get motivated toward digital literacy. The study by Gautam and Kanoujiya (2022) examined the impact of regional rural banks on digital literacy and rural development in India, using data from 29 Indian states and two union territories over three fiscal years. The study concluded that regional rural banks support digital literacy and rural development, and it advised banks and the government to concentrate on these issues to advance financial inclusion and rural development.

The article by Al Dahdah and Mishra (2022) examines India's transition to digital healthcare via the Rashtriya Swasthya Bima Yojana (RSBY) program and its use of smart cards. The authors discuss the politics of digitized public-private welfare policy and question the value of a program that aims to deliver affordable, high-quality healthcare to the private health market. The authors analyze digital access to healthcare in RSBY, questioning the role of digital technologies in transforming healthcare access in India. The study by Kameswaran *et al.* (2023) examines the challenges faced by people with visual impairments in India when accessing digital banking technology. The authors argue that there is a gap in research on the challenges faced by people with disabilities in obtaining accessible technology in the first place. Through qualitative research, the authors find that participants faced social and technical difficulties and engaged in advocacy work to secure and maintain access to digital banking. They expand on the view of advocacy as a form of access work performed by people with visual impairments.

The Department of Science and Technology (DST) has launched several pioneering initiatives in the realms of Data Science, Big Data, and the Internet of Things (IoT). These programs underscore the potential of data science in official statistics while also highlighting pertinent challenges (`https://dst.gov.in/data-science-research-initiative`, `https://dst.gov.in/big-data-initiative-1`, `https://dst.gov.in/internet-things-iot-research-initiative`). To enhance innovation policy delivery and monitoring in their respective sectors, DST also introduced the Automotive Sectoral System of Innovation (IASSI) and the Indian ICT Sectoral System of Innovation (IICTSSI) in 2023. (`https://dst.gov.in/sites/default/files/Indian%20Automotive%20Sectorial%20System%20of%20Innovation%20%28IASSI%29%20Report_0.pdf`, `https://dst.gov.in/sites/default/files/Indian%20ICT%20Sectorial%20System%20of%20Innovation%20%28IISSI%29%20Report_0.pdf`). Despite encountering challenges, both initiatives offer evidence-based development priorities and policy options, emphasizing the importance of effective management and connectivity for driving innovation and economic value.

4.1.2. Case study analysis

- Overview of official statistics in India:
  The MOSPI is in charge of the nation of Indian official statistics system. The system is responsible for gathering, putting together, and disseminating official statistics on different socio-economic indicators. The system consists of organizations, including NSSO, CSO, and RGCCI.

- Background on digital India program:
  Using technology to empower Indian society and economy, Digital India is a government initiative. Digital infrastructure, digital literacy, and digital services are intended to be created through the program. `https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1885962`). Through this program, citizens can access government services and information digitally. It focuses on infrastructure, governance, and services.

  The program includes initiatives such as the creation of digital infrastructure, such as the National Optical Fibre Network (NOFN) (`https://ddd.gov.in/scheme/bharat-net/`); the origination of digital literacy programs, the development of e-governance platforms and the promotion of digital financial services. Through this program, the government aims to empower citizens by providing them with digital tools and services that enhance their participation in the country's economic, social and political spheres.

- Evaluation of the implementation of digital India in official statistics:
  The implementation of digital India in official statistics has been ongoing since the launch of the initiative in 2015 (`https://csc.gov.in/digitalIndia`). Some key initiatives undertaken by the Ministry of Electronics and Information Technology (MeitY) under the Digital India program are Aadhaar, DigiLocker, Open Government Data Platform, etc. (`https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1885962`). The initiative has focused on several areas; including the following:

  – Digitization of data collection: The initiative has aimed to digitize data collection processes to improve the accuracy and timeliness of official statistics.
  – Development of digital platforms: The initiative has aimed to develop digital platforms for the dissemination of official statistics, such as the MOSPI website and mobile applications.
  – Use of data analytics: The initiative has aimed to leverage the power of data analytics to extract insights and patterns from official statistics.

Overall, the implementation of digital India in official statistics has led to several benefits; including the following:

- Improved accuracy and timeliness of official statistics: Digitalization has facilitated the speedy publication of official statistics and the reduction of errors.

- Increased accessibility of official statistics: The development of digital platforms has made official statistics more accessible to the general public, researchers, and policymakers.

- Increased efficiency of official statistics: The use of data analytics has improved the efficiency of official statistics analysis, allowing policymakers to make more informed decisions based on them.

However, the implementation of digital India in official statistics has also faced several challenges, including the following:

- Quality of data: The quality of data collected through digital platforms may be affected by problems such as incomplete or inaccurate data or bias in the sampling process.
- Privacy concerns: The digitization of data collection processes raises concerns about the privacy and confidentiality of individual data.
- Infrastructure challenges: The implementation of digital India in official statistics requires significant investment in digital infrastructure, which may be a challenge for some regions.

India's government uses advanced tools for data cleaning, standardization, and validation to improve data quality. For instance, the MOSPI has established the National Data Quality Forum (NDQF) to improve data quality across government agencies (`https://ndqf.in/`). The NDQF has implemented data quality scorecards and audits to ensure the accuracy and reliability of data. By spotting mistakes and abnormalities in data sets, the application of ML algorithms for predictive modeling has also improved the quality of data (Ngiam and Khor, 2019; Gruson *et al.*, 2019; Sharma *et al.*, 2020).

The MOSPI unveiled the digital India Mobile Van (`https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1895957`). Mobile Vans are unique initiatives under the program that provide digital literacy and allow remote and inaccessible areas in the country to access digital services. The vans are equipped with computers and other digital accessories, such as printers and scanners, which are used to provide several digital services. It includes digital services for connecting to the Internet, imparting digital literacy, and rendering government services electronically, such as enabling individuals to use the Internet for registering their birth and death certificates, etc. In addition to reaching out to women, seniors, and people with disabilities with limited access to digital infrastructure, the initiative ensures that digital services reach out to the most vulnerable communities. This initiative offsets the digital divide and ensures that marginalized groups access digital services from their residences, making it more inclusive and empowering citizens in different parts of the country. It has a significant contribution to ensuring the success of the Digital India program since all citizens have access to digital services irrespective of their geographical location or economic status.

On the other hand, the implementation of digital technologies in official statistics comes with various challenges, like data privacy and security issues. To solve this, the government of India has undertaken several measures. Firstly, the Data Protection Bills have been set (`https://www.meity.gov.in/writereaddata/files/The%20Digital%20Personal%20Data%20Protection%20Bill%2C%202022.pdf`). This bill primarily regulates how people's data can be collected, stored, and processed in India. Secondly, the National

Cyber Security Coordinator (NCSC) has been set to coordinate and oversight cybersecurity activities in government (https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1556474).

In conclusion, the execution of digital technologies in official statistics through initiatives such as Digital India has significantly improved the efficiency and effectiveness of government decision-making. It is still necessary to address quality issues and privacy concerns to ensure the correct and responsible use of official statistics. The Indian government has taken various measures to address these challenges, and continued efforts in this direction will be crucial for the success of digital government and official statistics in India.

4.2. Case study 2: Use of data science and BDA in the Indian census

The Indian census is one of the world's major administrative tasks, with over 1.39 billion people residing in India (https://statisticstimes.com/demographics/country/india-population.php). To make informed decisions, particularly in the healthcare, education, and infrastructure sectors, policymakers and government officials need the census. The application of data science and BDA has become more and more necessary for the analysis of the massive volumes of data collected during the census. This case study will explore the use of data science and BDA in the Indian census, focusing on their benefits and challenges.

4.2.1. Background

India has a census system that gathers socioeconomic and demographic data from each home every ten years. Businesses, researchers, and policymakers benefit from the census's valuable data. The introduction of digital technology has made the census more accurate and efficient than it was under the paper-based system. Under the British Raj, India conducted its first census in 1872. (https://censusindia.gov.in/nada/index.php/catalog/40444/download/44078/DROP_IN_ARTICLE-05.pdf). The Office of the Registrar General and Census Commissioner of India (ORGI), which is in charge of compiling and disseminating census data, conducts the census. (https://censusindia.gov.in/census.website/node/378).

4.2.2. Use of data science

With the tools it has provided for data collection, processing, analysis, and dissemination, data science has been crucial in the census. The census of India has been using data science and machine learning algorithms to improve the accuracy and efficiency of its operations. An example is the use of ML algorithms to improve the quality of data collection. The census uses Paper Data Capture Operation (PDC), which includes an automated data collection system that uses optical character recognition (OCR) to read the data collected from paper forms (https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/PDC_detailed_operational_plan.pdf). Additionally, the system alerts any data error such as missing or inaccurate entries for inspection by using ML methods. The gathered information becomes more accurate and trustworthy as a result.

4.2.3. Use of BDA

The article by Chatfield *et al.* (2018) focuses on the state of big data and BDA use in the National census context of four countries: Australia, Ireland, Mexico, and the U.S.A. The study found that the census agencies in these countries are at varying stages in digitally transforming their census process, products, and services through assimilating and using big data and BDA. However, the cross-case analysis of government websites and documents revealed emerging challenges in creating public value in the national census context, including BDA capability development, cross-agency data access & integration, and data security, privacy, and trust. Based on the insights gained, the article proposes a research model to explore the possible links among these challenges, BDA use, and public value creation.

The study by Marathe *et al.* (2020) presents a data science pipeline that integrates techniques such as ML, Statistics, Data Visualization, and Geographic Information System (GIS) for open big data in sustainable development. Using this pipeline, the Pune Municipal Corporation applied the geo-enabled tree census dataset to its tree census data. The study focuses on the visualization of big data, ward-wise analysis, and identification of marginalized species that require urgent attention from the authorities. A new biodiversity index is introduced in this study to address the limitations of existing indices when applied to cities in the Indian subcontinent. Overall, this study highlights the potential of data science techniques in analyzing big data and providing insights into sustainable development.

There are no studies utilizing BDA for the Indian census currently. BDA can be used in the Indian census to analyze large volumes of data and extract insights and patterns. By collaborating with technology companies, such as IBM and Microsoft, ORGI can develop BDA tools for the census. In addition to helping identify population characteristics like age, gender, education, and occupation, these tools will help analyze census data. As a result, policymakers and planners can gain a better understanding of demographic trends and patterns.

4.2.4. Challenges

Despite the advantages of data science and BDA in the Indian census, some issues still need to be resolved. One of the main challenges is data privacy and security. The census collects sensitive personal information, and there is a risk of this data being misused/breached. The ORGI has enforced strict protocols for regulating data handling and storage.

A significant number of Indians still lack access to digital technology due to the digital divide. Census data for these populations may be inaccurate or underrepresented as a result. Using offline methods and training field enumerators to collect data using non-digital means has been one of the strategies used by the ORGI to reach out to these populations.

4.2.5. Conclusion

Data science and BDA can improve data quality, accuracy, and efficiency in the Indian census. However, several challenges are addressed, including data privacy and security and the digital divide. To overcome these obstacles and guarantee that the census data is accurate, dependable, and secure, the ORGI should develop strategies and protocols. It is essential to balance technological innovation and ethical considerations when using census

data to ensure the responsible use of advanced technologies.

4.3. Case study 3: Artificial intelligence and official statistics in India

The National Sample Survey (NSS) data is used as an example in this case study to examine the application of AI in Indian government statistics. AI has the power to completely change the methods used to gather, handle, and evaluate official statistics. The study offers insights into the application of AI in the NSS data and looks at the advantages and difficulties of utilizing it in official statistics.

4.3.1. Background

Official data have always been gathered and published by the Indian government, going back to the colonial era. On the other hand, there is increasing interest in investigating the application of AI in official statistics due to the quick evolution of technology. AI can make data collection and analysis more quick, accurate, and efficient. This case study especially looks at the NSS data as an example of how AI is being used in official statistics in India. With a broad scope of social and economic variables covered, the NSS is the biggest household survey carried out in India. The NSS is India's largest household survey, covering socioeconomic indicators used by policymakers, researchers, and companies to understand its socioeconomic situation. Interviews and self-completed questionnaires are used for data collection. Traditional data collection and analysis are time-consuming.

AI is increasingly used in official statistics in India to improve data analysis, prediction, and decision-making. The study by Chawla *et al.* (2022) examines the role of AI and Information Management (IM) in India's energy transition, which has been strained due to rapid urbanization and modernization. Despite India's status as the global IT heart and having above-average research output in AI, it has not fully leveraged its benefits in the energy sector. The study analyzes proposed strategies, current policies, and available literature to highlight the challenges and barriers to developing and using AI and IM in India's energy sector. The study suggests that policymakers in India must take adaptive and swift actions toward developing comprehensive AI and IM policies to extract maximum benefits from the ongoing transition of the energy sector.

The article by Chatterjee *et al.* (2022) explores the public value generated by AI-enabled services from the perspective of Indian citizens. An analysis of 315 interviews is conducted using the Partial Least Square-Structural Equation Modeling (PLS-SEM) technique based on IT assimilation theory and public value theory. The study finds that the assimilation of AI-enabled services positively impacts citizens' satisfaction and generates public value. It also identifies risk factors that may influence the uptake of such services. The paper contributes to understanding the benefits and challenges of AI-enabled services in the public sector. For instance, the government has started several initiatives that use AI and ML algorithms to improve the timeliness and accuracy of official statistics. One such example is the use of chatbots for data collection and analysis. India introduces a WhatsApp chatbot to spread knowledge about the coronavirus and request social media platforms to stop the spread of false information. (`https://techcrunch.com/2020/03/21/india-whatsapp-mygov-corona-helpdesk-bot/`) The chatbot responds immediately to user inquiries, speeding up response times and increasing data accuracy.

Another example is the application of predictive modeling to estimate population growth and migration patterns. The study by Devi *et al.* (2022) analyzes the Land Use and Land Cover (LULC) change rate of Cochin, an urbanized coastal city in India. A contrast of the observed and simulated LULCs of 2020 validated the model's simulation. The model demonstrated acceptable LULC dynamics, with an overall accuracy of 87.5%. The future scenarios of LULC, projected till 2100, show an increase in built-up lands and a shrinkage of natural land covers, such as forests and water bodies. The urban growth indicator confirms the extreme transformation of the area in terms of urbanization. The study suggests establishing appropriate urban planning and management policies for sustainable environmental conservation.

In addition to the census, the government has also launched several AI-based initiatives to improve the collection and analysis of data in various sectors, including health, agriculture, and education. For instance, the National Health Stack (NHS) is a government initiative that aims to digitize health records and use AI to analyze the data to improve healthcare services. (`https://abdm.gov.in:8081/uploads/NHS_Strategy_and_Approach_1_89e2dd8f87.pdf`).

However, some challenges are involved. The study by Sharma *et al.* (2022) explores the interrelationships and challenges of implementing AI in India's Public Manufacturing Sector (PMS). AI integration with PMSs is challenging due to low data quality, inadequate understanding of cognitive technologies, privacy concerns, and the high cost of implementing cognitive projects. The study proposes a model for decision-makers and managers to develop intelligent AI-enabled systems for manufacturing organizations in emerging economies. The study highlights the need to address these challenges to enhance the scope of AI implementation in the PMS sector.

4.3.2. Methodology

This case study employs a qualitative research methodology, focusing on the study of secondary data sources. Based on literature, official reports, and interviews with field experts, the study examines the use of AI in NSS data. The analysis is guided by the following research questions:

- What are the benefits of using AI in official statistics, specifically in the NSS data?
- What are the challenges associated with implementing AI in official statistics?
- How has AI been implemented in the NSS data, and what are the implications of this implementation?

4.3.3. Findings

According to the study, AI has the potential to greatly increase the timeliness, accuracy, and efficiency of data gathering and processing in official statistics. AI can automate complex tasks like data imputation, validation, and cleaning. It can enhance data quality and lower mistake rates. Large data sets may be rapidly analyzed by AI, which enables researchers to identify patterns and insights that would be difficult to discover manually.

Moreover, AI can assist in lowering the expenses related to data collection and processing by replacing human labor.

Even with these benefits, there are still difficulties in integrating AI into official data. The quality of the data is one of the main obstacles. Unless AI algorithms are trained on accurate or neutral data, they will provide biased or inaccurate results. Data security and privacy are other issues since AI needs access to a lot of personal information. In low-resource environments, it might be difficult to find qualified workers to develop, deploy, and maintain AI systems.

In the case of the NSS data, AI has been implemented in several ways, such as using Natural Language Processing (NLP) techniques to extract data from open-ended survey questions and using ML algorithms to impute missing data. However, there are still challenges associated with the implementation of AI in the NSS data, such as the need for more training data to improve the accuracy of the algorithms.

4.3.4. Conclusion

In conclusion, the use of AI in official statistics has the potential to completely transform India's data collection, processing, and analysis methods. Incorporating AI into the statistics system is a positive step for the government. Addressing AI's issues and worries is essential, especially those about data security and privacy. AI has the potential to be a helpful tool in official statistics and advance national development with the correct policies and approaches.

5. Future directions and concluding remarks

There are several potentials for the future of official statistics in India as the benefits of data science, AI, and BDA continue to develop. When these cutting-edge technologies are combined, data gathering, processing, and distribution may become more accurate and efficient. Predictive modeling and ML may also help anticipate future trends and patterns, which can give policymakers and decision-makers important information. There are advantages and drawbacks, such as concerns about data privacy and quality, the digital divide, and ethics.

Digital government has a critical role in shaping the future of official statistics in India. It is possible to improve the accuracy, efficiency, and accessibility of official statistics by using advanced technologies. The digital government may also help different government departments and stakeholders engaged in official statistics collaborate and coordinate. It can also ensure that the benefits of official statistics are shared equitably among all sections of society, including marginalized communities. However, the digital government must also ensure that the ethical considerations associated with the advantages of advanced technologies are addressed and that the benefits of official statistics are shared equitably.

Collaboration among policymakers, researchers, and practitioners is essential for maximizing the benefits and minimizing the challenges of advanced technologies. Some recommendations for each group include:

- Policymakers: Policymakers must prioritize investment in technology and infrastruc-

ture to support the implementation of advanced technologies in official statistics. Additionally, these technologies must address ethical considerations, data quality, and privacy concerns. Furthermore, they must prioritize capacity building and training to ensure that government officials have the necessary skills and knowledge to implement these technologies effectively.

- Researchers: Researchers need to investigate the possible advantages and challenges associated with the implementation of advanced technologies in official statistics. In addition, they must develop and share best practices for the responsible use of these technologies. Furthermore, they must collaborate with government agencies to ensure that research findings are translated into policy and practice.
- Practitioners: Practitioners involved in official statistics must prioritize the development of data quality and management frameworks to ensure that data is accurate, reliable, and timely. The benefits of official statistics must also be shared equitably across society, including marginalized groups. Furthermore, they must engage in ongoing professional development to ensure they have the necessary skills and knowledge to implement advanced technologies effectively.

To sum up, the use of advanced technologies, such as data science, AI, and BDA, has the potential to transform the field of official statistics in India. Cooperation between researchers, policymakers, and practitioners will maximize benefits while minimizing obstacles. Digital governments must address the ethical issues associated with these technologies and ensure that all sections of society can share the benefits of official statistics.

In conclusion, this review article explored the use of data science, AI, and BDA in official statistics in India's digital government. This study found that the implementation of digital government programs has greatly enhanced the gathering, processing, and distribution of official statistics in India. While AI has aided in the development of predictive modeling and pattern recognition, data science and BDA have made it possible to collect and analyze data more thoroughly and accurately. However, there are several difficulties with using these technologies in official statistics, such as bias, privacy, and data quality concerns. To guarantee that the use of cutting-edge technology in official statistics is morally acceptable, responsible, and accurate, policymakers and practitioners need to address these issues.

The implications of this study are significant for both practice and policy. It emphasizes the importance of implementing cutting-edge technologies in official statistics to improve data quality and facilitate data analysis. Also, it emphasizes the importance of thoroughly considering ethical and privacy concerns before introducing new technologies. Digital government can promote the application of cutting-edge technology in official statistics, and government support is essential for advancing pertinent infrastructure and expertise. While using these technologies, policymakers should ensure that official statistics remain accurate and unbiased.

The future of official statistics in India depends on the development of digital government and the use of advanced technologies, such as data science, AI, and BDA. Future research in this area should explore the ethical and privacy concerns surrounding the benefit of these technologies in official statistics. Additionally, research can explore how these technologies can improve the quality and accuracy of official statistics in India's rural and

regional regions. Furthermore, research should examine how advanced technologies can facilitate the accessibility and dissemination of official statistics to decision-makers, researchers, and the general public.

Acknowledgements

The authors wish to express sincere gratitude to the Editors for their valuable guidance and counsel. We are deeply thankful to the anonymous referees for their constructive comments and insightful suggestions on the earlier version of this manuscript.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Abbas, S. W., Hamid, M., Alkanhel, R., and Abdallah, H. A. (2023). Official statistics and big data processing with artificial intelligence: Capacity indicators for public sector organizations. *Systems*, **11**, 424.

Al Dahdah, M. and Mishra, R. K. (2022). Digital health for all: The turn to digitized healthcare in India. *Social Science & Medicine*, **319**, 114968.

Alvarenga, A., Matos, F., Godina, R., and CO Matias, J. (2020). Digital transformation and knowledge management in the public sector. *Sustainability*, **12**, 5824.

Ashofteh, A. and Bravo, J. M. (2021). Data science training for official statistics: A new scientific paradigm of information and knowledge development in national statistical systems. *Statistical Journal of the IAOS*, **37**, 771–789.

Chatfield, A. T., Ojo, A., Puron-Cid, G., and Reddick, C. G. (2018). Census big data analytics use: International cross case analysis. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pages 1–10.

Chatterjee, S. (2020). AI strategy of India: policy framework, adoption challenges and actions for government. *Transforming Government: People, Process and Policy*, **14**, 757–775.

Chatterjee, S., Khorana, S., and Kizgin, H. (2022). Harnessing the potential of artificial intelligence to foster citizens' satisfaction: An empirical study on India. *Government Information Quarterly*, **39**, 101621.

Chawla, Y., Shimpo, F., and Sokołowski, M. M. (2022). Artificial intelligence and information management in the energy transition of India: lessons from the global IT heart. *Digital Policy, Regulation and Governance*, **24**, 17–29.

Chinnaswamy, A., Papa, A., Dezi, L., and Mattiacci, A. (2019). Big data visualisation, geographic information systems and decision making in healthcare management. *Management Decision*, **57**, 1937–1959.

Daraio, C., Leo, S. D., and Scannapieco, M. (2022). Accounting for quality in data integration systems: a completeness-aware integration approach. *Scientometrics*, **127**, 1465–1490.

Devi, A. B., Deka, D., Aneesh, T. D., Srinivas, R., and Nair, A. M. (2022). Predictive modelling of land use land cover dynamics for a tropical coastal urban city in Kerala, India. *Arabian Journal of Geosciences*, **15**, 399.

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., and Papadopoulos, T. (2019a). Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. *British Journal of Management*, **30**, 341–361.

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., Roubaud, D., and Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, **226**, 107599.

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., and Roubaud, D. (2019b). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change*, **144**, 534–545.

Gautam, R. S. and Kanoujiya, J. (2022). Role of regional rural banks in rural development and its influences on digital literacy in India. *Iconic Research and Engineering Journals*, **5**, 92–101.

Gautam, R. S., Rastogi, S., Rawal, A., Bhimavarapu, V. M., Kanoujiya, J., and Rastogi, S. (2022). Financial technology and its impact on digital literacy in India: Using poverty as a moderating variable. *Journal of Risk and Financial Management*, **15**, 311.

Ghosh, J. K., Maiti, P., Rao, T. J., and Sinha, B. K. (1999). Evolution of statistics in India. *International Statistical Review/Revue Internationale de Statistique*, **67**, 13–34.

Gokhberg, L., Kuznetsova, T., Abdrakhmanova, G., Fursov, K., Nechaeva, E., Shahsnov, S., and Suslov, A. (2020). Prospective model of official statistics for the digital age. *Higher School of Economics Research Paper No. WP BRP*, **111**.

Gootzen, Y., Daas, P. J. H., and van Delden, A. (2023). Quality framework for combining survey, administrative and big data for official statistics. *Statistical Journal of the IAOS*, **39**, 439–446.

Gruson, D., Helleputte, T., Rousseau, P., and Gruson, D. (2019). Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. *Clinical Biochemistry*, **69**, 1–7.

Guntukula, R. (2020). Assessing the impact of climate change on Indian agriculture: Evidence from major crop yields. *Journal of Public Affairs*, **20**, e2040.

Haim, M., Leiner, D., and Hase, V. (2023). Integrating data donations in online surveys. *Medien & Kommunikationswissenschaft*, **71**, 130–137.

Hantrais, L., Allin, P., Kritikos, M., Sogomonjan, M., Anand, P. B., Livingstone, S., Williams, M., and Innes, M. (2021). Covid-19 and the digital revolution. *Contemporary Social Science*, **16**, 256–270.

Hassani, H. and MacFeely, S. (2023). Driving excellence in official statistics: Unleashing the potential of comprehensive digital data governance. *Big Data and Cognitive Computing*, **7**, 134.

Ilyas, Q. M., Ahmad, M., and Mehmood, A. (2023). Automated estimation of crop yield using artificial intelligence and remote sensing technologies. *Bioengineering*, **10**, 125.

Jaber, M. M., Ali, M. H., Abd, S. K., Jassim, M. M., Alkhayyat, A., Aziz, H. W., and Alkhuwaylidee, A. R. (2022). Predicting climate factors based on big data analytics based agricultural disaster management. *Physics and Chemistry of the Earth, Parts A/B/C*, **128**, 103243.

Kameswaran, V., Y, V., and Marathe, M. (2023). Advocacy as access work: How people with visual impairments gain access to digital banking in India. *Proceedings of the ACM on Human-Computer Interaction*, **7**, 1–23.

Kasianova, N., Kendiuknov, O., and Pishenina, T. (2021). Features of the assessment of state and a prospective for the development of the digital economy. In *1st International Scientific Conference" Legal Regulation of the Digital Economy and Digital Relations: Problems and Prospects of Development"(LARDER 2020)*, pages 233–239. Atlantis Press.

Križman, I. and Tissot, B. (2022). Data governance frameworks for official statistics and the integration of alternative sources. *Statistical Journal of the IAOS*, **38**, 947–955.

Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., and Li, X. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, **26**, 234–252.

Malhotra, C., Anand, R., and Singh, S. (2018). Applying big data analytics in governance to achieve sustainable development goals (SDGs) in India. *Data Science Landscape: Towards Research Standards and Protocols*, **38**, 273–291.

Mamatha, K., Samantha, S., and Prasad, K. K. (2023). Crop yield prediction using deep learning. In *ICDSMLA 2021: Proceedings of the 3rd International Conference on Data Science, Machine Learning and Applications*, pages 93–102. Springer.

Marathe, A., Mirchandani, K., Chordiya, K., and Stephen, K. (2020). Big data analytics for sustainable cities: Pune tree census data exploratory analysis. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE.

Misra, N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., and Martynenko, A. (2020). IoT, big data, and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal*, **9**, 6305–6324.

Nair, V. (2021). Becoming data: biometric IDs and the individual in 'Digital India'. *Journal of the Royal Anthropological Institute*, **27**, 26–42.

Ngiam, K. Y. and Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, **20**, e262–e273.

Rana, N. P., Luthra, S., and Rao, H. R. (2020). Key challenges to digital financial services in emerging economies: the Indian context. *Information Technology & People*, **33**, 198–229.

Rao, T. J. (2010). Official statistics in India: The past and the present. *Journal of Official Statistics*, **26**, 215.

Rao, T. J. (2013). National statistical commission and Indian official statistics. *Resonance*, **18**, 1062–1072.

Rao, U. (2019). Population meets database: Aligning personal, documentary and digital identity in Aadhaar-enabled India. *South Asia: Journal of South Asian Studies*, **42**, 537–553.

Sakshaug, J. W. and Steorts, R. C. (2023). Recent advances in data integration. *Journal of Survey Statistics and Methodology*, **11**, 513–517.

Sarma, N. (1958). Economic development in India: The first and second five year plans. *Staff Papers-International Monetary Fund*, **6**, 180–238.

Saxena, S. (2018). Perception of corruption in e-government services post-launch of "Digital India": Role of demographic variables. *Digital Policy, Regulation and Governance*, **20**, 163–177.

Sharma, A., Jain, A., Gupta, P., and Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, **9**, 4843–4873.

Sharma, M., Luthra, S., Joshi, S., and Kumar, A. (2022). Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy. *Government Information Quarterly*, **39**, 101624.

Sinha, B. B. and Dhanalakshmi, R. (2022). Recent advancements and challenges of Internet of Things in smart agriculture: A survey. *Future Generation Computer Systems*, **126**, 169–184.

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. Z., Paul, R., Smriti, K., Naik, N., and Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science (1971-)*, **191**, 1473–1483.

Tantalaki, N., Souravlas, S., and Roumeliotis, M. (2019). Data-driven decision making in precision agriculture: The rise of big data in agricultural systems. *Journal of Agricultural & Food Information*, **20**, 344–380.

Tripathi, M. and Dungarwal, M. (2020). Digital India: Role in development. *International Journal of Home Science*, **6**, 388–92.

Vavilova, D. D. and Ketova, K. V. (2023). Information and analytical system for the analysis of regional socio-economic processes based on the integrated use of dynamic models of various types. *Journal Of Applied Informatics*, **18**, 97–110.

Vijai, C. (2019). FinTech in India–opportunities and challenges. *SAARJ Journal on Banking & Insurance Research (SJBIR)*, **8**.

Vijai, C. and Wisetsri, W. (2021). Rise of artificial intelligence in healthcare startups in India. *Advances in Management*, **14**, 48–52.