

# Literature Review on the Development and Applications of Data Science in Various Fields

Margaret

Politeknik Negeri Medan

---

## ARTICLE INFO

### Keywords:

Data Science, Applications, Big Data, Machine Learning, Artificial Intelligence

### Email:

[agetyedisa@gmail.com](mailto:agetyedisa@gmail.com)

---

## ABSTRACT

This study is a literature review aimed at describing the development and application of Data Science across various sectors of life. The method used involves a review of scientific literature from multiple academic sources published between 2018. The findings indicate that Data Science has evolved from classical statistical approaches to artificial intelligence-based systems that support decision-making in the health, finance, education, agriculture, industry, and government sectors. This review also highlights the integration of Big Data, Machine Learning, and Artificial Intelligence technologies as the main drivers of global digital transformation.

Copyright © 2025 JU-KOMI. All rights reserved are Licensed under a Creative Commons Attribution- NonCommercial 4.0 International License (CCBY-NC 4.0)

## INTRODUCTION

Data Science is an interdisciplinary field that combines statistics, computer science, and domain knowledge to extract insights from data. The advancement of digital technology and the abundance of data have made Data Science a key component in strategic decision-making processes across various sectors. In the past decade, terms such as Machine Learning, Artificial Intelligence, and Big Data have become inseparable parts of both research and industry.

The rapid development of information technology and computational capabilities over the past two decades has driven the emergence of an interdisciplinary discipline known as Data Science. Data Science integrates statistics, computer science, machine learning, and domain knowledge to extract insights from large-scale data (big data) and support evidence-based decision-making. This evolution is not merely an accumulation of statistical techniques, but a transformation of working paradigms—from descriptive analysis and business reporting to prediction, optimization, and automation based on complex data-driven models.

The applications of Data Science are now widespread across many sectors, including healthcare, finance, agriculture, education, manufacturing, energy, transportation, retail, and even smart governance (smart cities). Examples of its applications include AI-assisted disease diagnosis, financial fraud detection, agricultural weather prediction, personalized learning, predictive maintenance in industry, and route optimization in transportation. This transformation has also been driven by advancements in algorithms (e.g., deep learning), data infrastructure (cloud and distributed computing), and the availability of sensor/IoT data.

On the other hand, the adoption of Data Science raises several challenges, including data ethics, privacy concerns, model bias, data quality limitations, and the demand for skilled human resources. This literature review aims to summarize the historical development, methodological trends, and major applications of Data Science across various fields, while also identifying research gaps, ethical implications, and future directions. The specific

objectives of this review are: (1) to present a summary of the evolution and key drivers of Data Science development, and (2) to examine the main applications in selected fields and commonly used methods.

## METHOD

This review is structured as a narrative literature review that combines results from systematic searches with thematic synthesis. The methodological steps undertaken are as follows:

### Search Criteria

1. Keywords

“data science evolution,” “applications of data science,” “machine learning review,” “deep learning applications healthcare finance agriculture,” “data science ethics,” and their variations in both English and Indonesian.

2. Databases and Sources

This study utilizes secondary data from journal articles, conference proceedings, book chapters, and review reports from 2018 onward, along with several seminal earlier sources for historical context. Both primary studies and systematic reviews are included to provide a comprehensive overview.

### Selection and Screening

The initial screening was conducted based on titles and abstracts, followed by full-text reviews of candidate articles. The focus was on studies discussing methodological developments, real-world applications, evaluations, and ethical or implementation issues. Duplicate or irrelevant articles were excluded.

### Data Extraction

For each article, the following information was recorded: study objectives, application domain, employed methods (ML/DL algorithms, preprocessing techniques, infrastructure), datasets, main findings, limitations, and recommendations.

### Thematic Synthesis

Findings were organized thematically into:

1. evolution and enabling technologies;
2. sectoral applications (healthcare, finance, agriculture, manufacturing, transportation,
3. education, government/retail);
4. ethical issues, privacy, and implementation challenges; and future trends (digital twins, foundation models, multimodal integration, edge computing).

### Methodological Limitations

The narrative approach allows for broad synthesis but does not constitute a quantitative meta-analysis. Some recent literature may have been missed due to the rapid pace of publication. Future research is recommended to conduct domain-specific systematic reviews or quantitative meta-analyses.

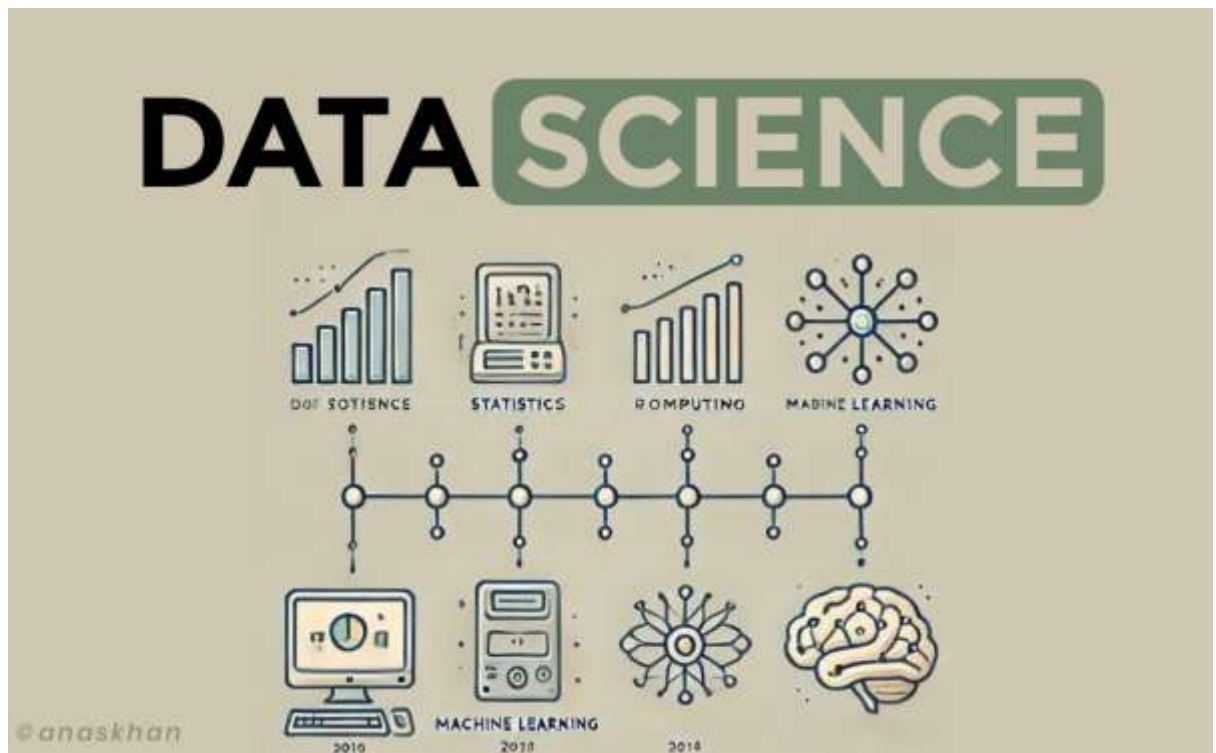
## RESULTS AND DISCUSSION

This section presents a summary of the main findings: the historical development and supporting technologies, followed by a discussion of sectoral applications, and concludes with key emerging issues.

## Historical Development and Supporting Technologies

### Brief History and Development Phases

The evolution of Data Science can be divided into several phases: the roots of statistics and probability (18th–19th century), the era of computation and computational statistics (1950–1990), the era of big data and distributed storage (2000s), and the era of machine learning and deep learning (2010s onward), characterized by large-scale models and end-to-end applications. This transformation has been driven by advancements in computational power, data storage capacity, and the availability of data from the internet, sensors, and digital transactions.



**Figure 1.** History of Data Science

Data science, a term that resonates across industries today, has a fascinating history that traces its evolution from simple statistical methods to the powerful and complex world of artificial intelligence (AI). What began as a way to interpret numbers has now become a dynamic field enabling groundbreaking innovations like self-driving cars, personalized recommendations, and predictive healthcare.

The journey of data science from simple statistics to complex AI highlights its transformative power. Its evolution has been fueled by technological advancements, curiosity, and the quest to make sense of the world through data. As data science continues to grow, it holds the potential to solve some of humanity's most pressing problems and unlock innovations

### Key Technologies

1. Infrastructure: Hadoop, Spark, cloud computing, and data lake/warehouse architectures enable the storage and processing of large-scale datasets.
2. Algorithms: Advances in Machine Learning algorithms (e.g., ensemble methods, SVM, gradient boosting) and breakthroughs in Deep Learning (e.g., CNN, RNN, Transformer)

have opened new applications, particularly in image processing, natural language understanding, and real-time signal analysis.

3. Tools & Ecosystem: The open-source ecosystem (such as scikit-learn, TensorFlow, and PyTorch) and MLOps platforms (such as Kubeflow and MLflow) have accelerated the research-to-production cycle.

### **Sectoral Applications: Key Findings and Examples**

#### a. Healthcare

Data Science is transforming medical practice from symptom-based diagnosis to data-driven diagnosis using radiological images, genomics, and electronic health records. Common applications include cancer detection from medical images using CNNs, clinical outcome prediction (readmission, mortality) through predictive modeling, genomic analysis for personalized therapy, and real-time patient monitoring via wearables. Numerous review studies demonstrate the effectiveness of AI models in improving diagnostic accuracy and operational efficiency. However, clinical evaluation, external validation, and regulatory compliance remain major challenges. Issues and challenges: limited labeled datasets, representation bias (non-representative of heterogeneous populations), model interpretability (explainability), and privacy regulations (HIPAA/GDPR and local equivalents). Clinical validation and integration into physicians' workflows are also critical needs.

#### b. Finance

In finance, Data Science facilitates fraud detection, credit risk assessment, portfolio optimization, and algorithmic trading. Supervised learning and anomaly detection techniques are widely used. ML models can identify anomalous transaction patterns faster than rule-based systems. However, challenges include adapting to rapidly evolving fraud patterns and ensuring model explainability for regulatory compliance.

#### c. Agriculture

Data Science supports precision agriculture by enabling yield prediction, irrigation optimization, pest detection via multispectral imagery, and fertilizer recommendation based on soil prediction models. The integration of satellite data, drones, soil sensors, and ML models enables data-driven decisions to enhance productivity and sustainability. Recent review studies highlight the promising role of deep learning in processing satellite and drone imagery.

#### d. Manufacturing & Energy (Industry 4.0)

Predictive maintenance using IoT sensors and predictive models reduces downtime, while process optimization through streaming data analytics enhances efficiency. The adoption of digital twins allows the simulation of physical assets for improved operational planning. These innovations increase productivity and reduce operational

costs. However, integrating legacy systems and achieving data standardization remain obstacles to large-scale implementation.

e. Transportation & Smart Cities

Applications include route optimization, traffic prediction, fleet management, and smart city planning. The combination of sensor data, telematics, and ML-based predictive models improves transport service responsiveness and safety. Challenges include real-time data scalability, interoperability across infrastructures, and user location privacy concerns.

f. Education and Social Research

2. In education, learning analytics supports personalized learning, early detection of students at risk of dropping out, and evaluation of curriculum effectiveness. In social research, Data Science enables public opinion analysis, modeling of information diffusion, and evidence-based policymaking. However, educational data use requires careful ethical consideration regarding student privacy and the responsible use of automated decision systems.

**Table 1.** Comparison of Data Science Methods and Applications Across Sectors

Sector	Common Data Science Methods	Main Objective	Example Applications	Technologies Used
Healthcare	Machine Learning, Deep Learning	Disease diagnosis, clinical outcome prediction	Cancer detection, diabetes risk prediction	Python, TensorFlow, Scikit-learn
Finance	Data Mining, Predictive Analytics	Risk analysis, fraud detection	Credit scoring, fraud detection	R, SAS, Power BI
Education	NLP, Clustering	Learning behavior analysis	Learning material recommendation, dropout prediction	Python, RapidMiner
Agriculture	IoT Data, Time Series Forecasting	Crop yield optimization	Weather prediction, plant disease detection	MATLAB, Google Cloud AI
Industry & Manufacturing	Predictive Maintenance, Computer Vision	Production efficiency	Product defect detection	TensorFlow, AWS ML
Government	Data Integration, Big Data Analytics	Data-driven policymaking	Smart city, crime prediction	Hadoop, Spark

This table illustrates the diversity of Data Science methods and applications across various sectors. Each field employs specific approaches depending on its objectives and data characteristics. For instance, the healthcare sector heavily relies on Deep Learning for medical image recognition, whereas the finance sector emphasizes data mining and predictive analytics for risk management.

#### Dominant Methods and Techniques

1. Supervised Learning: regression and classification (tree-based models, SVM, neural networks).
2. Unsupervised Learning: clustering (k-means, hierarchical), dimensionality reduction (PCA, t-SNE, UMAP).
3. Deep Learning: CNNs for image data, RNNs/Transformers for sequential and textual data. Transformers and large language models (LLMs) are now revolutionizing NLP and multimodal applications.
4. Ensemble Methods & Gradient Boosting (XGBoost, LightGBM): popular for tabular tasks (e.g., credit scoring).
5. MLOps & Deployment: data pipelines, model versioning, performance monitoring, and automated deployment have become essential practices for production environments.

**Table 2** Summary of Studies by Domain

Domain	Author(s) (Year)	Study Focus	Key Findings
Healthcare	Rane et al. (2024)	Applications of ML/DL in healthcare and other sectors	Deep Learning provides high accuracy but still requires clinical validation
Agriculture	Zhu et al. (2018)	Smart agriculture – DL and sensors	Deep Learning shows great promise in pest and harvest detection, but data availability remains limited
Finance	Oluwabiye et al. (2023)	Deep Learning in finance	Increasing publication trend, focusing on fraud detection and credit scoring
Evolution & General	Raban & Gordon (2020)	Evolution of Data Science and bibliometric trends in Big Data	Integration between Big Data and Data Science is increasingly important; literature has grown rapidly since ~2010
Education	Huang, Xin & Chang (2025)	Machine Learning in educational process analysis	High potential for personalized learning, though significant ethical challenges remain

#### Ethical, Privacy, and Regulatory Issues

As many studies have shown, the adoption of Data Science is not merely a matter of technology; it carries ethical (bias, fairness), legal (data protection), and social (job displacement, decision transparency) implications. Model interpretability (XAI – Explainable AI) is crucial in critical domains such as healthcare and finance. National and international

regulations, along with ethical guidelines, are essential to ensure that implementation delivers benefits without harming vulnerable groups.

### Research Gaps and Recommendations

Based on the synthesis of the literature, several research gaps and future directions are identified as follows:

1. External Model Validation: Many studies remain at the prototype stage without extensive validation across diverse populations—essential for ensuring generalizability.
2. Data and Representation: Imbalanced or non-representative datasets lead to bias—there is a need for diversified datasets and standardized audit protocols.
3. Explainability & Human-in-the-Loop: Combining AI systems with human expertise is crucial for final decision-making, particularly in critical domains.
4. MLOps and Reproducibility: Standardizing pipelines, promoting reproducible research practices, and implementing post-deployment monitoring are necessary for reliability.
5. Ethics & Regulation: Interdisciplinary research integrating legal, ethical, and technological perspectives is needed to develop comprehensive implementation guidelines.

### CONCLUSION

Multidimensional Transformation: Data Science has evolved from its statistical roots into an interdisciplinary discipline supported by advances in computing, ML/DL algorithms, and data infrastructure. This transformation has reshaped how organizations operate and make decisions across various sectors. Broad Applications and Tangible Impact: Evidence from the literature shows significant applications of Data Science in healthcare, finance, agriculture, manufacturing, transportation, education, and the public sector—each contributing to greater efficiency, accuracy, and service personalization. However, full-scale implementation requires validation, regulatory adaptation, and ethical consideration. Challenges and Research Needs: Dataset limitations, model bias, explainability requirements, and the lack of standardized MLOps practices remain major challenges. Future research should focus on external validation, methods for fairness and interpretability, reproducible practices, and ethical-regulatory studies. Future Directions: Emerging trends include the integration of large language and multimodal models, digital twins, edge AI for on-device processing, and the strengthening of data governance and policy practices. Cross-disciplinary collaboration—among technical, clinical, and policy domains—will be essential to maximize benefits while minimizing risks.

### REFERENCES

- Chinthamu, N., & Karukuri, M. (2023). *Data Science and Applications*. Journal of Data Science and Intelligent Systems, 1(2), 83–91. <https://doi.org/10.47852/bonviewJDSIS3202837>
- Ramachandran, K. K. (2023). *Data Science in the 21st Century: Evolution, Challenges and Future Directions*. Data Science – Extracting Insights, 1(1), 2–8. <https://doi.org/10.70705/ppp.dsei.2023.v01.i01.pp2-8>

- Priestley, J. L., & McGrath, R. J. (2019). The Evolution of Data Science: A New Mode of Knowledge Production. *International Journal of Knowledge Management*, 15(2), 97–109. <https://doi.org/10.4018/IJKM.2019040106>
- Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, 122, 1563–1581. <https://doi.org/10.1007/s11192-020-03371-2>
- Costa, C. J., & Aparicio, M. (2023). Applications of Data Science and Artificial Intelligence. *Applied Sciences*, 13(15), 9015. <https://doi.org/10.3390/app13159015>
- Tang, M. Z. (2023). The analytics and applications of big data and machine learning. *Highlights in Science, Engineering and Technology*, 31. <https://doi.org/10.54097/hset.v31i.4817>
- Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. *SBC Reviews on Computer Science*, 4(1), 28–43. <https://doi.org/10.5753/reviews.2024.3757>
- Zhu, N. Y., Liu, X., Liu, Z. Q., Hu, K., Wang, Y. K., Tan, J. L., ... Guo, Y. (2018). Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *International Journal of Agricultural & Biological Engineering*, 11(4), 32–44. <https://doi.org/10.25165/j.ijabe.20181104.4475>
- Huang, J., Xin, Y. P., & Chang, H. H. (2025). The Application of Machine Learning to Educational Process Data Analysis: A Systematic Review. *Education Sciences*, 15(7), 888. <https://doi.org/10.3390/educsci15070888>
- Rane, N. L., Mallick, S. K., Kaya, Ö., & Rane, J. (2024). Applications of machine learning in healthcare, finance, agriculture, retail, manufacturing, energy, and transportation: A review. In *Applied Machine Learning and Deep Learning: Architectures and Techniques* (pp. 112–131). Deep Science Publishing. [https://doi.org/10.70593/978-81-981271-4-3\\_6](https://doi.org/10.70593/978-81-981271-4-3_6)
- Alpaydin, E. (2021). *Introduction to Machine Learning* (4th ed.). MIT Press. DOI: <https://doi.org/10.7551/mitpress/11249.001.0001>
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media. DOI: <https://doi.org/10.5555/3570980>