

# Book Review:

## Encyclopedia of Data Science and Machine Learning (5 Volumes)

Leigh Wang, Northwestern University, USA\*

### ABSTRACT

The Encyclopedia of Data Science and Machine Learning (EDSML) examines current, state-of-the-art research in the areas of data science, ML, data mining (DM), optimization, artificial intelligence (AI), statistics, and the interactions, linkages, and applications of knowledge-based business with information systems. It provides an international forum for practitioners, educators, and researchers to advance the knowledge and practice of all facets of BDML, emphasizing emerging theories, principles, models, processes, and applications to inspire and circulate cutting-edge findings into research, business, and communities. This encyclopedia contains a collection of 187 high-quality chapters, which were written by an international team of more than 370 experts representing leading scientists and talented young scholars from more than 46 countries and regions.

### KEYWORDS

Artificial Intelligence, Data Mining, Data Science, Information Systems, Machine Learning, Optimization, Statistics

### INTRODUCTION

Big Data and Machine Learning (BDML) are driving and harnessing the power of the Fourth Industrial Revolution, also called Industry 4.0 or 4IR, which revolutionizes how companies, organizations, and institutions operate and develop. BDML has incessantly explored its “depth” and expanded its “breadth” in order to maximize the citizens’ “wealth” while promoting society’s “health”. The Encyclopedia of Data Science and Machine Learning (EDSML) examines current, state-of-the-art research in the areas of data science, ML, data mining (DM), optimization, artificial intelligence (AI), statistics, and the interactions, linkages, and applications of knowledge-based business with information systems. It provides an international forum for practitioners, educators, and researchers to advance the knowledge and practice of all facets of BDML, emphasizing emerging theories, principles, models, processes, and applications to inspire and circulate cutting-edge findings into research, business, and communities (Wang, 2023).

This encyclopedia contains a collection of 187 high-quality chapters, which were written by an international team of more than 370 experts representing leading scientists and talented young scholars from more than 46 countries and regions, including Algeria, Argentina, Austria, Bangladesh, Brazil, Canada, Chile, China, Colombia, Cuba, Denmark, Egypt, El Salvador, Finland, France, Germany, Ghana, Greece, Hong Kong, Hungary, Indonesia, Iran, Iraq, Japan, Lebanon, Macau, Malaysia,

DOI: 10.4018/IJBAN.319321

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Mexico, Netherland, New Zealand, Poland, Portugal, Saudi Arabia, Serbia, Singapore, South Africa, Sweden, Switzerland, Syria, Taiwan, Tunisia, Turkey, UK, USA, Venezuela, Vietnam, etc. They have contributed great effort to create a source of solid, practical information, informed by the sound underlying theory that should become a resource for all people involved in this dynamic new field.

Big Data has now become a critical part of the business world and daily life, as the synthesis and synergy of Machine Learning (ML) and Big Data (BD) have enormous potential. With its recent release in October 2022, this book is up-to-date with the latest DSML developments. This is a comprehensive resource on data science and machine learning. It has over 3,100 pages and covers a wide range of topics in the field, making it a valuable reference for students, researchers, and professionals alike.

Research in data science and ML is important because it propels the development of models, technologies, and methods for collecting, analyzing, and making sense of large amounts of data. These technologies and techniques have wide-ranging applications, from advancing scientific discovery, predicting customer behavior, and improving business decision-making and bottom lines, to optimizing manufacturing and service outcomes.

With the mounting amount of data being generated in various channels and domains, DSML has become increasingly crucial for organizations and individuals to extract meaningful insights and make fast, frequent, and fact-based decisions. Moreover, research in these areas also pushes the boundaries of what is possible and helps to create brand-new ideas, techniques, and algorithms for tackling increasingly complex problems, leading to better solutions, further innovations, and progress. Research in DSML can lead to significant advancements in various fields such as service, green energy, finance, healthcare, and transportation and positively bear on the entire society. Companies that embrace and implement advanced DSML techniques can expand their competitive advantage and drive growth.

There are several reasons why research in DSML can be challenging: Addressing human bias from ML models needs careful consideration; Extracting significant insights from heterogeneous, high-dimensional, and complex data is tough; Selecting the suitable model for a given problem can be a demanding task; Interpreting results is difficult, especially for deep neural networks. Obtaining labeled data can be difficult or expensive; Overfitting is quite common with ML models, where the model may suit the training data very well, but does not generalize well to unseen datasets; Keeping up with ever-evolving technologies is necessary. Keeping up with the advancements with the latest research and developments can be a challenging task.

By analyzing large amounts of data and discovering patterns and relationships, machine learning algorithms can be used to make predictions, automate tasks, and improve decision-making. DSML drives the development of technologies and methods for analyzing and making sense of large amounts of data. These technologies have a wide range of applications, from improving business decision-making and predicting customer behavior to advancing scientific discovery and optimizing healthcare outcomes.

With the growing amount of data being generated in various industries and domains, DSML has become increasingly crucial for organizations and individuals to extract meaningful insights and make informed decisions. Moreover, research in these areas also pushes the boundaries of what is possible and helps to create new solutions for complex problems, leading to further innovation and progress.

This comprehensive volume is a critical resource for BDML practitioners and scientists, including industry professionals, technical managers, and corporate executives. In addition, educators, scholars, graduate students, and government officers will also find this book to be useful. Among the perspectives examined include historical underpinnings, major breakthroughs, discoveries, authoritative research results, strategic planning and policy, hot job prospects, and future directions, all of which are focused on the unique benefits and capabilities found in Industry 4.0.

According to Mike and Hazzan (2023), the history of scientific paradigms can be divided into four major phases. The first scientific paradigm established thousands of years ago was based on empirical science, in which scientists observed and described natural phenomena. The second scientific

paradigm, applied hundreds of years ago, was the theoretical paradigm, in which scientists developed models to explain natural phenomena. The third scientific paradigm, introduced several decades ago, is the computational paradigm, in which scientists use algorithms and computers to simulate complex phenomena. The fourth scientific paradigm, as described by Gray, is data exploration, in which data is captured or simulated and then analyzed by scientists to infer new scientific knowledge. The National Institute of Standards and Technology (NIST) considers data science to be the current evolution of the fourth paradigm.

*The Encyclopedia of Data Science and Machine Learning* from IGI Global publisher has only a few non-direct competitors. Here is a list:

- “Handbook of Big Data Analytics” by Härdle et al.
- “The Oxford Handbook of Computational and Mathematical Psychology” by Busemeyer et al.
- “Machine Learning Algorithms” by Giuseppe Bonaccorso.
- ” Real-World Machine Learning” by Brink et al.

These books have focused on different areas with special scopes, so they complement rather than compete with *the Encyclopedia of Data Science and Machine Learning*.

After experiencing significant challenges and breakthroughs in BDML, I highly recommend that you persuade your institution to invest in this comprehensive volume for the benefit of you and your colleagues. I am confident that this resource will prove invaluable, and you will not regret the decision to purchase it.

## REFERENCES

- Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.
- Brink, H., Richards, J., & Fetherolf, M. (2016). *Real-world machine learning*. Simon and Schuster.
- Busemeyer, J. R., Wang, Z., Townsend, J. T., & Eidels, A. (2015). *The Oxford handbook of computational and mathematical psychology*. Oxford University Press. doi:10.1093/oxfordhb/9780199957996.001.0001
- Härdle, W., Lu, H. H. S., & Shen, X. (Eds.). (2018). *Handbook of big data analytics*. Springer International Publishing. doi:10.1007/978-3-319-18284-1
- Mike, K., & Hazzan, O. (2023). What is Data Science? *Communications of the ACM*, 66(2), 12–13. doi:10.1145/3575663
- Wang, J. (Ed.). (2023). *Encyclopedia of Data Science and Machine Learning* (Vols. 1–5). IGI Global., doi:10.4018/978-1-7998-9220-5