Journal of Data Science 23 (2), 449–450
April 2025
DOI: 10.6339/25-JDS232BRV
Book Review

Book Review: "Veridical Data Science: The Practice of Responsible Data Analysis and Decision Making" by Bin Yu and Rebecca L. Barter

Xizhi Wu[1]

[1]*School of Statistics, Renmin University of China, Beijing, China*

Veridical Data Science, authored by Bin Yu and Rebecca L. Barter, offers a pioneering and rigorous exploration of data science grounded in real-world problem-solving. Moving away from the "true model" framing and avoiding what Box (1976) termed "mathematistry" (theory for its own sake, divorced from practice), the book presents a refreshing approach that prioritizes practical relevance. Its distinctive methodology sets it apart from other works in the field of statistics and data science.

At a time when the discipline of statistics faces a crossroads (He et al., 2019) in its evolution, this book provides a clear direction for transformation. With the Predictability, Computability, and Stability (PCS) framework as its foundation, the authors systematically outline the data science life cycle (DSLC). This framework, developed through years of teaching and research, serves as a powerful tool for bridging theoretical concepts and their application in practice.

The book combines scientific rigor with engineering precision, presenting data science as a craft that demands both technical skill and philosophical depth. Its primary goal is to foster independent and critical thinking, guiding readers to learn not only data science techniques but also how to approach problems methodically. Readers from diverse backgrounds will find its concepts and practices highly accessible and inspiring.

The book is organized into three parts. Part I introduces the fundamentals of veridical data science, including an overview of the data science life cycle and practical advice on setting up data science projects. Part II focuses on preparing, exploring, and describing data, with topics ranging from data preparation and exploratory data analysis to principal component analysis and clustering. Part III addresses prediction, covering both theoretical and practical aspects of predictive modeling, including regression methods, decision trees, random forests, and the production of final prediction results. Each chapter is carefully structured to balance conceptual depth with practical application, providing a comprehensive guide for learners at all levels.

Rather than extensively covering machine learning algorithms, which are already well-documented in existing books and online resources, the authors focus on conceptual foundations and problem-solving strategies. This approach ensures that the book remains both unique and highly relevant to data scientists and statisticians seeking to strengthen their theoretical and practical understanding.

Importantly, *Veridical Data Science* is freely available online at https://vdsbook.com, making its insights widely accessible to both academic and professional audiences.

In summary, this book stands as an indispensable resource for those navigating the rapidly evolving landscape of data science. Its emphasis on methodological rigor and practical relevance ensures that readers are well-equipped to tackle contemporary data-driven challenges with confidence and clarity.

* Email: xizhi.wu@aliyun.com.

© 2025 The Author(s). Published by the School of Statistics and the Center for Applied Statistics, Renmin University of China. Open access article under the CC BY license.

References

Box GEP (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356): 791–799. https://doi.org/10.1080/01621459.1976.10480949

He X, Madigan D, Yu B, Wellner J (2019). Statistics at a crossroads: Who is for the challenge. *NSF workshop report*, National Science Foundation.