Book Review

Reviewed by Zhongxian Wang and Sylvain Jaume*

Email: wangj@montclair.edu
Email: sjaume@saintpeters.edu
*Corresponding author

Mining Your Own Business: A Primer for Executives on Understanding and Employing Data Mining and Predictive Analytics
by: Jeff Deal and Gerhard Pilcher
Published 2016
by Data Science Publishing,
300 West Main St., Suite 301, Charlottesville, VA 22903,
USA, 161pp
ISBN: 978-0-9967121-0-1

With the Age of Big Data upon us, should we risk being drowned in a struggling cyclone of digital data? Or should we instead be lifted to an ever higher level of knowledge by capitalising on a reservoir of meaningful flowing information? Facing the explosion of big-data and the crucial need to acquire insight into it using specialised software, organisations need to increase their operational efficiency and gain competitive advantage by making use of this potential treasure trove of valuable information (Wang, 2014).

Data mining is a new frontier for many modern businesses. Knowledge discovery in multiple databases offers many opportunities and challenges (Adhikari et al., 2017). However, there is a big gap between theory and practice. On the one hand, data mining and predictive analytics have been demonstrated as the 21st century's most powerful new management tools; on the other hand, these high-level tools are too complicated to many high-ranking executives. With many vivid real-world examples, Deal and Pilcher fill the gap by sharing their extensive experience over decades. The authors go into great detail to elaborate as to what this new frontier upholds in the modern business world.

There are four sections. Section 1 of 'Introduction and overview' covers two chapters. The authors notice the knowledge explosion in the modern times in Chapter 1 of 'Empowering the decision makers'. Data mining and predictive analytics help decision makers filter raw data and focus it on key issues. We should not be scared because data analytics does not replace human reasoning; it intensely leverages it.

Chapter 2 deals with 'Clearing up the confusion'. The authors cite Fast and Elder's framework and emphasise that data analytics techniques can be organised into the following ten levels:

1 Standard and ad hoc reporting.

2 Statistical analysis.

3 Unsupervised learning.

4 Business rules and alerts.

5 Simulation.

6 Optimisation.

7 Parameter learning.

8 Structure learning.

9 Ensembles.

10 Casual modelling.

This list elaborates how data mining can truly be applied to the initial business practices of a business step by step. Advanced analytics (Levels 5–10) typically offer the greatest return on investment.

Section 2 of 'The analytic organisation' consists of four chapters. In Chapter 3 of 'Leading a data analytics initiative', the authors state that obtaining buy-in early on from fellow stakeholder is key because implementing a data analytics initiative and changing a corporate culture is never easy. Also, a small initial success serves as the foundation for greater success. In Chapter 4 of 'Staffing a data analytics project', the authors emphasise the importance of identifying a visionary when building a data analytics team. The team should not be afraid to ask questions that may appear foolish and is willing to try different approaches, even at the risk of short-term failure.

Chapter 5 focuses on 'Acquiring the right tools'. Having the people with the right knowledge and experience is more important than having the right tools. As Chai and Shih (2017) pinpoint that there is almost a blind faith that sophisticated algorithms can be used to explore huge databases and find interesting relationships independent of any theories or prior beliefs. Fancy tools only cannot make an organisation analytically competent. Also, open-source tools are a popular and growing segment of the analytic domain. The high-level programming languages R and Python are especially sophisticated and widely used. In line with the authors expressed in Chapter 6 of 'Hiring data analytics consultants', jumping into data analytics without professional assistance often lead to trouble or even failure. Most organisations need substantial help over a long time period of launching and implementing data analytics projects. Subject-matter experts (SMEs) need to be on call for the analytics team, but their commitment level is usually modest and affordable.

Section 3 of 'The modelling process' is the core of the entire book with five chapters. Starting from Chapter 7 of 'Understanding the data mining process', the authors clarify that a data analytics initiative is an iterative process. Lessons learned at each step, as a feedback, can prompt more focused questions that can lead to a re-examination and a re-execution of prior steps. Chapter 8 highlights 'Understanding the business'. Data scientists need to understand how the business functions and how it will use the data. The business questions the model is to answer should be carefully figured out. If separated from the main goals of the business, even the most advanced model will produce inconsequential or distorted results.

Chapter 9 focusing on 'Understanding and preparing the data'. Accurate data is vital for an analytics project. Proficient data scientists know how to work with messy data,

they have skills to get around most data problems. They need to build data dictionaries and talk with data experts in the organisation to understand what the data fields represent. Data miners should invite data owners to participate in its early stages of an entire process in order to promote buy-in and collaboration on the project.

According to the authors in Chapter 10 of 'Building the model', no single model type will work for every type of business problem. Each model has its strengths and weaknesses as the current research demonstrates. Over-fitting is a common mistake due to 'the law of diminishing returns'. The goal of data scientists is not to arrive at the best *technical* solution, but at best *practical* solution, as well as keeping in minds the needs, constraints, and resources of the client. Overlooking both validations could cause serious problems later as Chapter 11 illustrations. Technical validation checks how well the model performs on new data whereas business validation checks how well the model satisfies its business purpose. The authors suggest to use Elder's 'The top 10 data mining mistakes' and target shuffling to achieve statistical significance.

The last two chapters make Section 4 of 'Putting the model into practice'. According to the authors in Chapter 12 of 'Deploying the model', the final step of model deployment should be an integral part of the initial planning due to its significant impact on budgeting, personnel, and institutional processes. Periodically back-testing all models with current data and conditions should not be ignored. Management must proactively focus on potential, but not reactively focus on past events. Also, the authors encourage a culture that fosters innovation and tolerates the risk of failure in Chapter 13 of 'Realising the transformation'. A tipping point may not appear very soon. Perseverance is critical.

This new frontier helps analyse and give us vivid visions of patterns in a business in order to improve the business practices that are already set in place. Before applying useful tools to the process it is important to understand the data mining process and how crucial it is to understand how to use the data. The chart in the book on page 68 explains this entire process to the very core of the data. It starts with the understanding of the business and data followed by preparation of data and modelling, and continues onto evaluation and deployment of proven theory with the data. One of the most important factors mentioned in this part of the book is the need for experts who can interpret your models and put them to use in order to justify your findings.

Data is only useful if it can be interpreted in the proper way through subject-matter expertise. Although the data does not have to be perfect in order to be useful, the authors stress this throughout the later chapters, the right form of the data and the right volume of the data must be presented at the proper time. This book is a practical guide for organisational leaders and top-level executives. If you want to get solid advice and learn instructive cases of data mining, this is the one book you must read.

References

Adhikari, A., Jain, L.C. and Prasad, B. (2017) 'A state-of-the-art review of knowledge discovery in multiple databases', *Journal of Intelligent Systems*, Vol. 26, No. 1, pp.23–34.

Chai, S. and Shih, W. (2017) 'Why big data isn't enough', *MIT Sloan Management Review*, Vol. 58, No. 2, pp.57–61.

Wang, J. (Ed.) (2014) *Encyclopedia of Business Analytics and Optimization*, 1st ed., IGIGlobal, Hershey, PA, USA.