

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,700

Open access books available

180,000

International authors and editors

195M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Introductory Chapter: Overview of Data and Decision Sciences – Recent Advances and Applications

Tien M. Nguyen

1. Introduction

Due to the rapid growth of big data analytics related to all aspects of human activities, the surge in decision-making complexity due to the current climate of uncertainty with unforeseen consequences, and the increasing pervasiveness of advanced information and communication technologies (ICT) such as the proliferation of mobile applications, Internet-of-Things, and bots, we have witnessed an acceleration of integration of many complex ICT systems-of-systems (SoS) and social networks across of a wide spectrum of application domains that include, but are not limited to, telecommunications, satellite communications, medicine, military, education, agriculture, arts, and culture. The primary motivation for this book is to compile some of the latest research work addressing recent advances and applications of data and decision sciences (DDS) across the above-mentioned application domains. This book is a collective effort that uses a diverse set of studies and investigations to cover a wide spectrum of DDS applications. The goal is to shed some insights into the use of DDS models for assisting data analysts and decision-makers.

The objective of this introductory chapter is two-fold, namely, to provide (i) an overview of the data science and decision science, (ii) recent advances and DSS applications with an emphasis on machine learning and artificial intelligence (ML-AI), and (iii) overview and understanding of recent DDS applications. The remaining of this chapter is organized as follows:

- Section 2 provides an overview of data science and decision science.
- Section 3 discusses the differences between data science and decision science and recent advances in DDS.
- Section 4 concludes the chapter with final remarks on the DDS trends.

2. Overview data science, and decision science

2.1 Data science

Data science is a relatively new and emerging field of research for many mathematicians, statisticians, scientists, and engineers in the world. It has been

derived from data mining along with statistical analysis. It is defined in Cambridge Dictionary as “the use of scientific methods to obtain useful information from computer data, especially large amounts of data [1]”. In a more technical detail definition in Dictionary.com, it is defined as a field that “deals with advanced data analytics and modeling, using mathematics, statistics, programming, and machine learning to extract valuable, often predictive information from large data sets [2]”. Practically, IBM defines data science as a science field, which combines mathematics and statistics, specialized programming languages, advanced data analytics, artificial intelligence (AI), and machine learning (ML) with specific subject matter expertise to uncover actionable insights hidden in an organization’s data. These insights can be used to guide decision-making and strategic planning [3]. A data science life cycle used by the industry is captured on the home page of the University of California in Berkley, School of Information [4]. The data science life cycle includes five stages, namely, (i) Stage 1 – data capture stage: data acquisition, data entry, signal reception, and data extraction; (ii) Stage 2 – data maintenance stage: data warehousing, data cleansing, data staging, data processing, data architecture; (iii) Stage 3: data mining processing stage: data mining, clustering/classification, data modeling, data summarization; (iv) Stage 4 – data analysis stage: exploratory/confirmatory, predictive analysis, regression, text mining, qualitative analysis; and (v) Stage 5 – communication stage: data reporting, data visualization, business intelligence, decision making. In the context of this book, this Subsection 2.1 focuses on IBM’s definition and data processing and analysis stages of the data science life cycle, including data mining, machine learning and artificial intelligence (ML-AI) using neural networks and deep learning, statistical learning, and Bayesian statistics.

2.1.1 Data mining

Big data analytics (BDA) is defined as the process of exploiting and extracting meaningful information from a large and complex collection of data¹. Data mining is one of the key required functions in the BDA process. It’s well-known that the BDA process is part of the data science¹ life cycle, including five data processing stages. As pointed out earlier, the data processing Stage 3 is the data mining processing (DMP) to discover data patterns from a large collection of data [5–8]. **Figure 1** illustrates the DMP characteristics, including the types of data that can be mined and analyzed, the kinds of data mining patterns, data mining techniques, and applications.

As shown in **Figure 1**, DMP can be performed on various types of data such as (i) data from databases collected past and current banking data or experimental data from a complex satellite system; (ii) data from data warehouses, e.g., Amazon data warehouse contained mass data from Amazon business transactions collected from multiple sources and stored in a unified schema; (iii) actual daily real-time transaction data from the banks, e.g., credit approvals, check approvals, payment approvals, etc.; and (iv) other type of data includes but not limited to data collected from information technology (IT) system, data collected from health care and medical sciences, data collected from military defense systems such as images, video data streams, etc. For the kinds of data mining patterns, one of the key components of the data mining patterns is the characterization and discrimination of the features of a target class of data objects against the other features of objects from one or multiple contrasting target classes. After characterization and discrimination processing, the

¹ <https://www.bmc.com/blogs/big-data-vs-analytics/>.

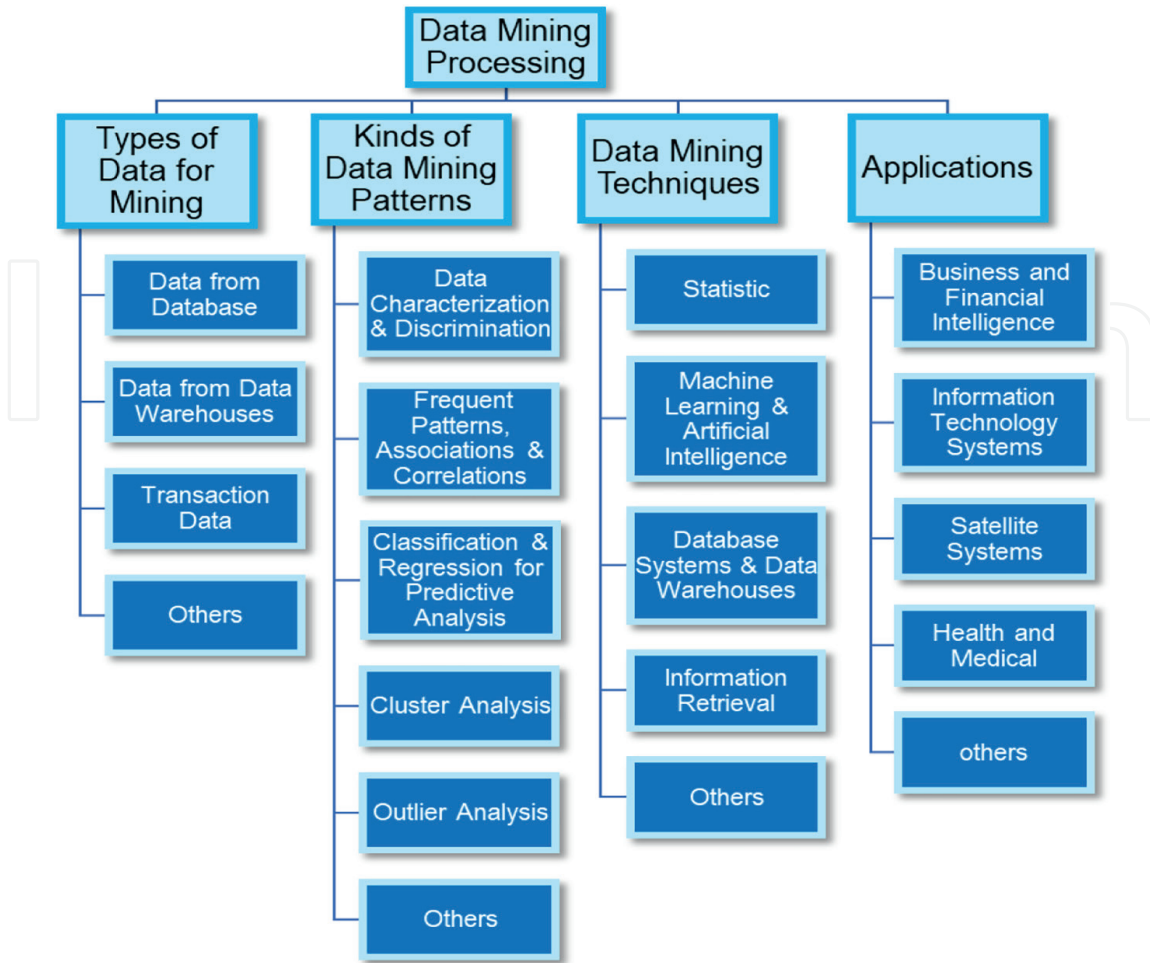


Figure 1.
 Data mining processing (DMP) characteristics.

data is analyzed for frequent patterns using association and statistical correlation analyses. An example of the frequent pattern analysis is to study the behavior of the computer consumers in terms of how often and the types of the computer they buy, the software they buy, and the buyers’ profession and their income ranges.

Within the context of this chapter, the following subsections will discuss ML-AI and statistical methods of interest to the data mining techniques as shown in **Figure 1**.

2.1.2 ML-AI, statistical learning, and Bayesian statistics

ML-AI techniques are usually used for estimating and predicting the data characteristics and associated data trends. For examples, ML-AI can be used to analyze big data to predict stock market trends [9], and analysis of big consumer data can help the suppliers to forecast trends of customer behavior, markets, prices, and so on [10]. When the data content has several categorical variables, the prediction can be achieved through classification and pattern recognition. As an example, ML-AI using supervised learning and support vector machine (SVM) can be used to (i) predict the impacts of signal distortions caused by non-ideal satellite operational environment to the transmitted signal components, and (ii) classify the source of signal distortions [11]. In terms of ML-AI, past data is used to train the system, thus the newly accumulated data represents the case of repeat “modeling” where new data will be used to predict the trend in the future or classify an object (e.g., signal component) to a group

(e.g., source of signal distortion) by comparing with the old data. Majority of practical and useful ML-AI modeling techniques are usually stochastic or statical in nature. Therefore, the term statistical learning is also used in literature for ML-AI modeling. As pointed out in [12], Bayesian is a way of practicing statistics in which the ML-AI modeling is built upon probability distributions, i.e., the modeling is solely calibrating and adjusting the probabilities. Thus, Bayesian statistics utilize Bayes theorem and facilitate the calculation of posterior distributions as follows:

$$p(\theta / data) \propto p(data / \theta) \times p(\theta) \quad (1)$$

Where $p(\theta / data)$ is the posterior probability distribution, $p(data / \theta)$ is the likely hood or the classification probability, and $p(\theta)$ is the prior probability.

Bayesian functional data analytical techniques include multiple curve-fitting (MCF), single neuronal analysis (SNA), and population-level analysis (PLA) [12]. MCF uses hierarchical modeling of firing intensity curves using BARS (i.e., BAR chart) approach. SNA is used for testing equality of two or more curves. Finally, PLA is used for testing equality of two groups of curves. As examples, MCF, SNA, and PLA have been used in health care and biology applications [13–15], respectively.

2.1.3 ML-AI using neural network and deep learning

ML-AI modeling uses neural networks (NN) and deep learning (DL) to model the neuronal cells and their intricate functionality, and networking for processing the data (i.e., information) [12]. The terminology NN-DL or deep NN (DNN) is usually used to indicate a network that has more than two hidden layers with an input layer and one output layer with multiple nodes, as shown in **Figure 2** [12, 16]. The variables $x_i, w_i^{(m)}$, and y_j are the DNN's parameters that are defined as the input node, weight of the hidden layer node, and the output node, respectively. As pointed out in [16] DNN model has more hidden layers, which requires longer simulation time and more training data storage.

The DNN modeling requires (i) characterizing the DNN's system parameters and the associated “loss” function in terms of the weight parameter $w_i^{(m)}$, and (ii) tune these parameters using the training data collected by the system architect under controlled environment. **Figure 3** provides a high-level description of the key DNN tuning parameters, including layer size and related mini-batch size for numerical approximation of gradient, gradient threshold, and learning rate. **Figure 3(a)** illustrates the layer size; **Figure 3(b)** shows the exploding gradient if the “terms” in the differential equation are greater than 1, and **Figure 3(c)** depicts the learning rate. **Figure 3(c)** shows that the learning rate can have a large learning rate and a small learning rate that can be used for fast adaptation during data acquisition phase and slow adaptation during tracking phase after the loss function is converged. Note that a differential equation is usually used to characterize a neural network layer.

In practice for DNN, there are usually four hyper-parameters to tune, namely, layer size, mini-batch size, gradient threshold, and learning rate. Tuning the layer size to select the best size to produce the best manageable agent size of training data meaning that the layer size should be selected to optimize the required training

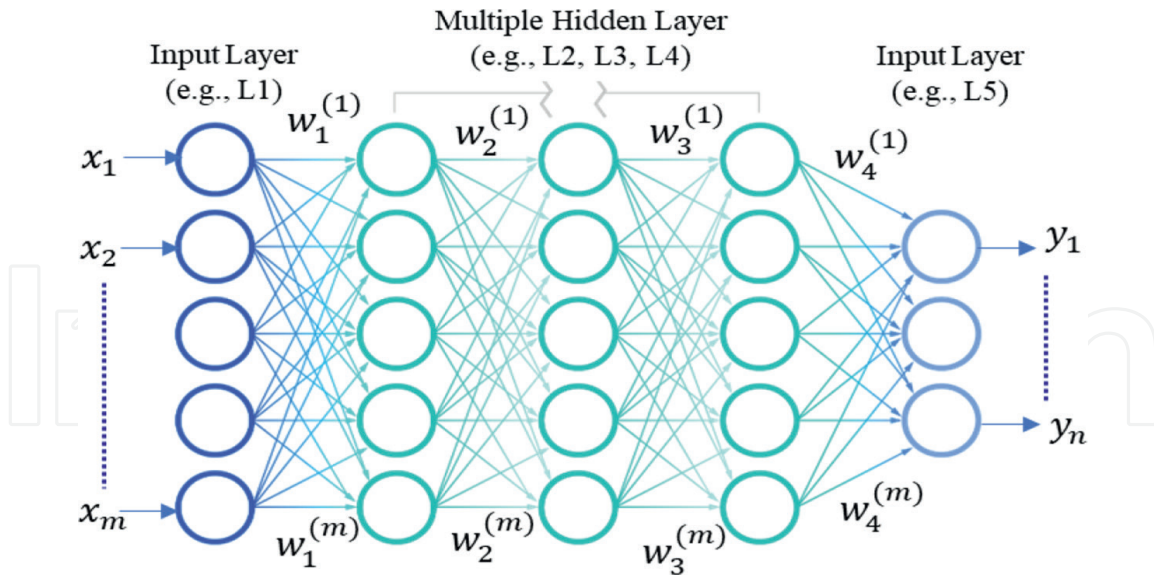


Figure 2.
 Deep neural network.

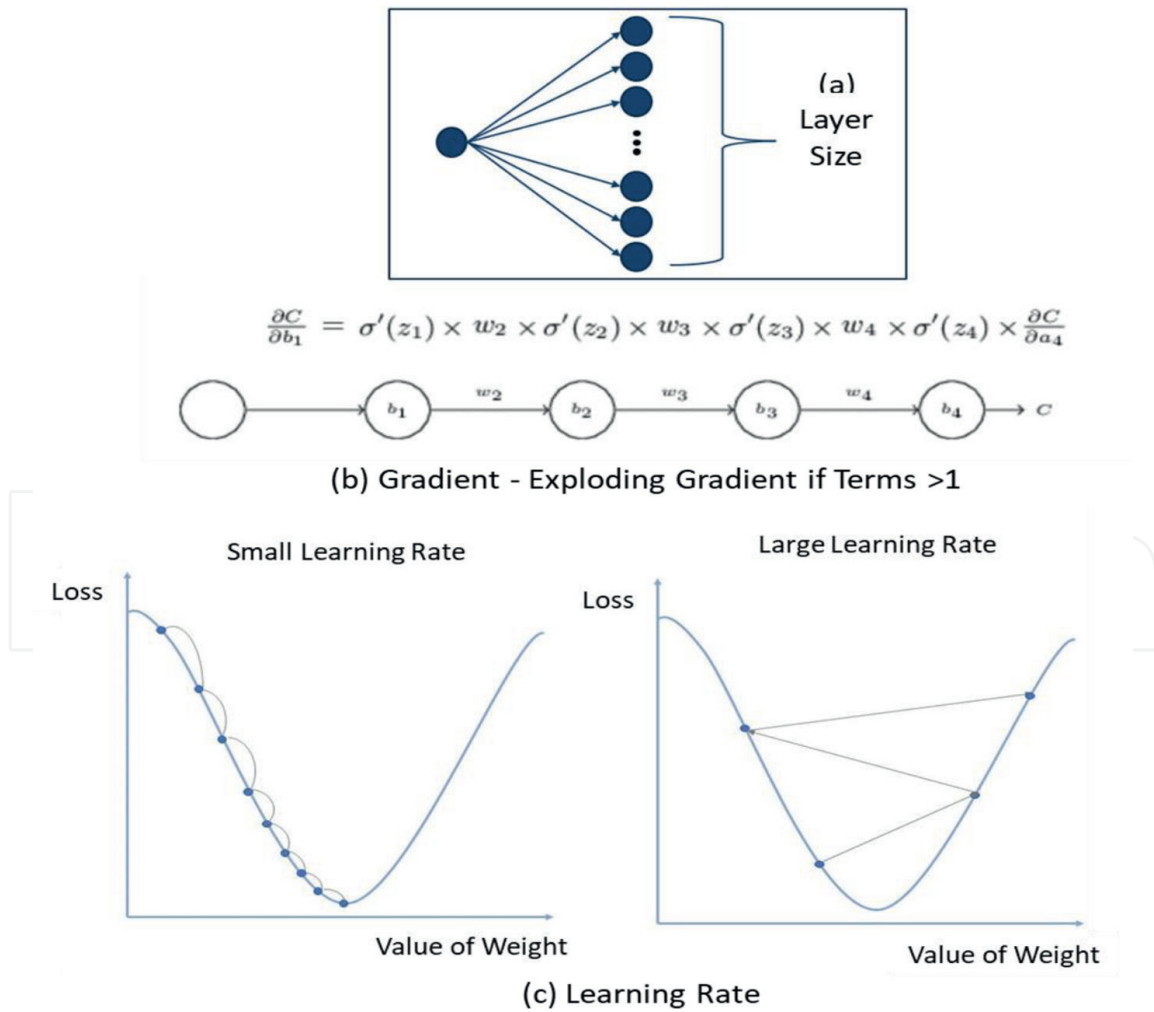


Figure 3.
 Deep neural network (DNN) and associated tuning parameters.

data storage. Tuning the mini-batch size to get the best size for the numerical approximation of the gradient. Tuning the gradient threshold to obtain the best gradient clipping to avoid an “exploding gradient” and the best step size to achieve a timely gradient descent or ascent step. Finally, the tuning of the learning rate is required to achieve the best reward/stopping criteria and learning rate criteria for better convergence. Ref. [16] describes the tuning process for an application of DNN in the design and development of future global navigation and satellite system (GNSS).

2.1.4 AI and expert systems

Earliest example of rule-based expert system was DENDRAL a system for identifying chemical structures developed in the 1960s at Stanford University [17]. DENDRAL was the first system that was called AI and expert system because the decision-making process and problem-solving behavior of organic chemistry were automated to identify unknown organic molecules. Since then, many systems were derived from DENDRAL including MYCIN, REX, MOLGEN, PROSPECTOR, XCON, STEAMER, etc. As an example, MYCIN system was developed in the 1970s to help physicians diagnose meningitis and bacterial infections [18]. As another example, REX system was developed in the 1980s and it was written with the language LIPS from Bell Labs. REX system had advanced the AI and expert system by incorporating rule-based guidance for simple linear regression. The name REX was derived from Regression EXpert, and it was an interface between humans (or users) and statistical software, and an interactive modeling software (IMS). The IMS was created to allow the user interacts with the statistical software more effectively [19].

Since then, the AI and expert systems have undergone rapid evolution. Especially, the COVID pandemic had stimulated private companies to invest in smart and advanced technologies using machine learning and AI, expert systems, cloud computing, and the Internet of Things (IoT) that enable their businesses to make better, more informed decisions in the presence of uncertain environment and fast-changing conditions. As pointed out in [20], currently, ML-AI and expert systems are typically designed and built for specific applications to address specific business or organization needs or technical challenges. They can be classified into two categories, namely, (i) forward chaining ML-AI Expert System (FC/ML-AI-ES) that uses data to predict future events, and (ii) backward chaining ML-AI Expert System (BC/ML-AI-ES) that uses historical data to understand why something occurred. Examples of FC/ML-AI-ES are forecasting inventory demand, or future crop conditions associated with specific geographic areas, etc. Examples of BC/ML-AI-ES are medical diagnostics or troubleshooting complex technical issues in hardware and software systems. A typical ML-AI-ES consists of three primary components, namely, knowledge base (KB), inference engine (IE), and user interface (UI). KB is defined as the data that the ML-AI-ES uses and works with. Modernized KB has automated capabilities that can organize the data and present the data as the user requested it (a.k.a. curate). IE is defined as part of the ML-AI-ES that applies logical rules and related mathematical and/or simulation models (a.k.a. algorithms) that can pull intelligent insights from KB based on user queries. Finally, UI is defined as the means through which a user interacts with KB through a commercial off the shelves software (COTS) platform. **Figure 4** depicts a high-level architecture of a ML-AI-ES [20].

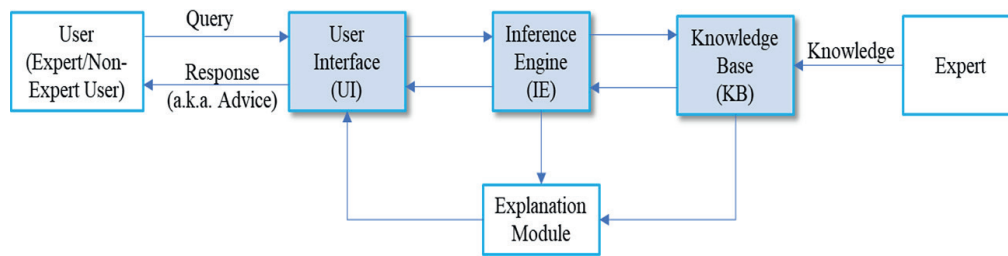


Figure 4.
Typical high-level ML-AI-ES architecture.

2.2 Decision science

Unlike data science, the root of decision science has been found in open literature dated back in the 1930s with an application to economic [21]. As defined by Harvard Chan School of Public Health, Decision Science is the collection of quantitative techniques used to inform decision-making at the individual and population levels². It includes decision analysis, risk analysis, cost–benefit and cost-effectiveness analysis, constrained optimization, simulation modeling, and behavioral decision theory, as well as parts of operations research, microeconomics, statistical inference, management control, cognitive and social psychology, and computer science. With the emergence of ML-AI and digital technologies, decision science ranges from traditional decision theories and analysis to advanced decision theories using emerging decision optimization techniques leveraging game theory, ML-AI, and ML-AI combined with mathematical modeling and simulation (M&S) techniques.

Basically, the traditional decision theory and analysis deal with the reasoning that drives a person's decision, or organization's choice, or a business's decision. In general, the traditional decision theory and analysis consist of three core concepts, including (i) elicitation and interpretation of the decision maker's preferences, (ii) the search of available options, and (iii) the management of uncertainty, risks, and regrets [21–23]. For a large organizations or collective settings involved multiple options associated with different users' needs and interests, the decision-making process is extended to multiple stakeholders. In the 1950s, Von Stackelberg, Nobel laureate John Nash, and Von Neumann are universally credited for their pioneering work on using game theory applied to decision-making process [24–26]. They proposed mathematical models of strategic interaction among rational decision-makers. The latter can be either cooperative or non-cooperative.

3. Data and decision sciences (DDS): Recent advances on DDS

As discussed in the previous sections, data and decision sciences (DDS) are inter-related. Data science involves with data collection, data mining, and data analysis. While decision science involves with the process of making decisions through interpretation of the data. But the data interpretation requires data analysis that is a subset of data science. The data analysis is usually conducted by applying mathematical and simulation models and related algorithms for optimizing the risks associated with the decision-making process.

² <https://chds.hsph.harvard.edu/approaches/what-is-decision-science/>.

3.1 Recent advances in data and decision sciences

As the Industry Revolution 4.0³ evolves to 5.0, the decision-making process is being challenged by massive data sources and the digitization of the business world along with the rise of environmental uncertainty and risks. Emerging big data analytics, ML-AI, and digitization technologies have allowed for seamless integration of data and decision sciences (DDS). The decision support system (DSS) using big data analytics and ML-AI approach is one of the recent advancements in the integration of DSS (a.k.a. advanced DSS). Along with big data analytics and ML-AI technologies, advanced DSS is a computer-based system that allows for the digitization of the decision-making processes using sophisticated mathematical and simulation models, and advanced optimization techniques. The DSS is designed to allow the decision-makers to make either optimal or satisfactory decisions in the presence of uncertainties. As an example, recently, a group of researchers at Aerospace Corporation has collaborated with North Carolina State University (NCSU), the University of Hawaii at Manoa (UH Manoa), and California State University in Fullerton (CSUF) to develop an advanced DSS tool supporting the development of the optimum acquisition strategy for buying complex space systems with optimum cost and acquisition risk [27–29]. The developed DSS tool leverages multi-criteria decision analysis process, game theory, and advanced optimization techniques to determine the best space system architecture solution and corresponding optimum acquisition strategy to acquire (a.k.a. buy) the system of interest. Finally, as indicated in the table of content, this book has collected a set of chapters addressing the recent advancement of DDS in space, business, medical, and agriculture applications.

4. Conclusion

This chapter has provided an overview of the data science and decision science (DDS) and discussed recent advances and DSS applications with a focus on ML-AI technology and its technology enablers. The introductory chapter complements this book's technical content by addressing other DDS topics and applications that are not presented in the chapters presented in this book. It should be pointed out that these book chapters share a common thread of DDS with topics ranges from recent DDS advancements to optimization modeling in decision science and cognitive decision-making process. For each DDS topic, the chapters provide an excellent introduction and background of the DDS problems making these chapters reachable to various scientific and engineering disciplines. Furthermore, for each chapter, the author also takes an effort to (i) discuss technical details associated with the proposed DDS models, and (ii) provide examples to demonstrate the use of the models. This effort from the authors will make their chapters to reach a wide range of readers across many scientific and engineering fields.

We hope that the readers find in this introductory chapter along with book chapters provide intriguing concepts and ideas that would help them solve their DDS problems and a source of ideas for their own work.

³ Industry Revolution 4.0 involves with digitization for automation using cyber physical systems on connected devices, and big data analytics. Industry Revolution 5.0 involves with mass customization and personalization for humans using cognitive computing and human intelligence.

IntechOpen

Author details


Tien M. Nguyen^{1,2}

1 California State University in Fullerton (CSUF), California, USA

2 The Aerospace Corporation, USA

*Address all correspondence to: tien.m.nguyen@aero.org

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Definition of Data Science. Available from: <https://dictionary.cambridge.org/us/dictionary/english/data-science>
- [2] Definition of Data Science. Available from: <https://www.dictionary.com/browse/data-science>
- [3] Data Science. Available from: <https://www.ibm.com/topics/data-science>
- [4] What's Data Science. Available from: <https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>
- [5] Han J, Kamber M, Pei J. *Data Mining Concepts and Techniques*. MA, USA: Elsevier Inc, Morgan Kaufmann Publishers; 2012. Available from: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- [6] Ramageri BM. Data mining techniques and applications. *Indian Journal of Computer Science and Engineering*. 2010;1(4):301-305
- [7] Han J. Data mining: Concepts and techniques. In: Department of Computer Science University of Illinois at Urbana-Champaign, chapters 6, Lecture Notes, January 20, 2018. ©2006 Jiawei Han and Micheline Kamber, all rights reserved. Available from: <https://www3.cs.stonybrook.edu/~cse634/ch6book.pdf>
- [8] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. *AI Magazine Volume 17 Number American Association for Artificial Intelligence*, © AAAI, 1996. Available from: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230/1131>
- [9] Jamal H. Using Big Data to Predict Stock Market Trends. 2023. Available from: <https://medium.datadriveninvestor.com/using-big-data-to-predict-stock-market-trends-312751189d76>
- [10] Seyedan M, Mafakheri F. Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data*. 2020;7-53:1-22, Springer Open. DOI: 10.1186/s40537-020-00329-2
- [11] Mendez-Villanueva J, Sopena G, Nguyen TM, Lee CH, Chen Y, Behseta S, et al. Innovative multicarrier broadband waveforms classification using machine learning for future GNSS applications. In: *SPACEOPS 2023 Conference Proceedings, the 17th International Conference on Space Operations*, 6-10 March 2023, Dubai, United Arab Emirates. Virginia, USA: AIAA
- [12] Behseta S. Introduction to Neural Networks and Deep Learning, Lecture Notes Part 1 and Part 2, California State University in Fullerton, Center for Computational and Applied Mathematics (CCAM), 2022. Available from: sbehseta@fullerton.edu
- [13] Cunningham JP, Gilja V, Ryu SI, Shenoy KV. Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Networks*. 2009;22(9):1235-1246. DOI: 10.1016/j.neunet.2009.02.004
- [14] Gupta P, Balasubramaniam N, Chang H-Y, Tseng F-G, Santra TS. A single-neuron: Current trends and future prospects. *Cells*. 2020;9:1528. DOI: 10.3390/cells9061528. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7349798/pdf/cells-09-01528.pdf>

- [15] Xu J, Chen G, et al. Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome Biology*. 2019;**20**:243. DOI: 10.1186/s13059-019-1859-0
- [16] Nguyen TM, Aguilar J, Lee CH, Paniagua-Rodriguez D, Shen D, Chen G, et al. Onboard HPA pre-distorter using machine learning and artificial intelligence for future GNSS applications. In: 2023 SPIE Conference Proceedings, April 30-May 4, 2023, Orlando, Florida. Bellingham, Washington, USA: SPIE
- [17] DENDRAL - Computer Software Expert System. Available from: <https://en.wikipedia.org/wiki/Dendral>
- [18] MYCIN - Backward Chaining Expert System. Available from: <https://en.wikipedia.org/wiki/Mycin>
- [19] Pregibon D, Gale WA. REX: An expert system for regression analysis. In: Proceedings of the 1984 International Association for Statistical Computing, Vienna, Italy. Washington, USA: American Statistical Association; 1984. pp. 242-248
- [20] Grant M. AI and Expert Systems: Powering the Future of Business, LeanIX. LeanIX Blogposts. 2022. Available from: <https://www.leanix.net/en/blog/artificial-intelligence-expert-systems>
- [21] Zeuthen F. On the determinateness of the utility function. *Review Economic Studies*, Oxford University Press. 1937;**4**(3):236-239
- [22] Simon H. The Sciences of the Artificial. Massachusetts: MIT Press; 1969
- [23] Schlaifer R. Analysis of Decisions under Uncertainty. New York: McGraw-Hill; 1969
- [24] Von Stackelberg HF. Market Structure and Equilibrium. 1st edition (translation into English) ed. Vol. 2011. Bazin: Urch & Hill: Springer; 1934
- [25] Nash JF. The bargaining problem. In *Econometrica*. Econometric Society. 1950;**18**(2):155-162
- [26] Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. New Jersey: Princeton University Press; 1953
- [27] Nguyen TM, Guillen A, Matsunaga S, Tran HT, Bui TX. “War-Gaming Applications for Achieving Optimum Acquisition of Future Space Systems,” a Book Chapter in the book titled “Simulation and Gaming,” London, UK: INTECH-Open Science-Open Minds; 2018. DOI: 10.5772/intechopen.69391
- [28] Nguyen TM, Tran HT, Guillen AT, Bui TX, Matsunaga SS. Acquisition war-gaming technique for acquiring future complex systems: Modeling and simulation results for cost plus incentive fee contract. *Journal of Mathematics*, MDPI - Publisher of Open Access Journals. 2018;**6**(3):1-29. DOI: 10.3390/math6030043. Available from: www.mdpi.com/journal/mathematics
- [29] Nguyen TM, Freeze T, Bui TX, Guillen A. Multi-criteria decision theory for enterprise architecture risk assessment: theory, modeling and results. In: Proc. SPIE 11422, Sensors and Systems for Space Applications. Vol. XIII. 2020. p. 114220. DOI: 10.1117/12.2559317