Encyclopedia of Data Science and Machine

John Wang (Montclair State University, USA)

Acknowledgment

The editor would like to thank all authors for their insights and excellent contributions to this major volume. I also want to thank the many anonymous reviewers who assisted me in the peer-reviewing process and provided comprehensive and indispensable inputs that improved our book significantly. In particular, the Editorial Advisory Board members, including Xueqi Cheng (Chinese Academy of Science), Verena Kantere (University of Ottawa, Canada), Srikanta Patnaik (SOA University, India), Hongming Wang (Harvard University), and Yanchang Zhao (CSIRO, Australia), have all made immense contributions in terms of advice and assistance, enhancing the quality of this volume. My sincere appreciation also goes to Prof. Alexander Tuzhilin (New York University). Despite his busy schedule, he has written three forewords for my consecutive encyclopaedias, over an 18-year span, in this expanding and exploring area.

In addition, the editor wishes to acknowledge the help of all involved in the development process of this book, without whose support the project could not have been satisfactorily completed. I owe my thanks to the staff at IGI Global, whose support and contributions have been invaluable throughout the entire process, from inception to final publication. Special thanks go to Gianna Walker, Angelina Olivas, Katelyn McLoughlin, and Melissa Wagner, who continuously prodded me via email to keep the project on schedule, and to Jan Travers and Lindsay Wertman, whose enthusiasm motivated me to accept their invitation to take on this project.

I would also like to extend my thanks to my brothers Zhengxian, Shubert (an artist, https://portraitartist.com/detail/6467), and sister Joyce Mu, who stood solidly behind me and contributed in their own unique ways. We are thankful for the scholarships which we have been provided, without which it would not have been possible for all of us to come and study in the U.S.

Finally, I want to thank my family: my parents for supporting me spiritually throughout my life and providing endless encouragement; my wife Hongyu for taking care of two active and rebellious teenagers, conducting all family chores, and not complaining to me too much.

This book was special due to the stresses and difficulties posed by the Covid-19 pandemic. We thank and salute the authors who had to overcome numerous challenges to help make this volume a reality. Our authors had to persevere through unprecedented circumstances to enable this masterful encyclopedia. Now, it is time to celebrate and reap the fruits of our demanding work! Cheese and cheers!

Preface

Big Data and Machine Learning (BDML) are driving and harnessing the power of the Fourth Industrial Revolution, also referred to as Industry 4.0 or 4IR, which revolutionizes the way companies, organizations, and institutions operate and develop. With the age of Big Data upon us, we risk drowning in a flood of digital data. 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.

BDML not only deals with descriptive and predictive analytics but also focuses on prescriptive analytics through digital technology and interconnectivity. It has continuously explored its "depth" and expanded its "breadth". BDML will remain to maximize the citizens' "wealth" while promoting society's "health".

The *Encyclopedia of Data Science and Machine Learning* 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, 2022).

How can a manager get out of a data-flooded "mire"? How can a confused decision-maker navigate through a "maze"? How can an over-burdened problem solver clean up a "mess"? How can an exhausted scientist bypass a "myth"? The answer to all of the above is to employ BDML.

As Roy et al. (2022) point out, data has become the center point for almost every organization. For quite a long time, we are familiar with Descriptive Analytics (what happened in the past) and Diagnostic Analytics (why something happened in the past), as well as Predictive Analytics (what is most likely to happen in the future). However, BDML could go much above and beyond them with Prescriptive Analytics (what should be done now), which recommends actions companies, and organizations can take to affect those outcomes. The digital transformation, the horizontal and vertical integration of these production systems, as well as the exploitation via optimization models, can make a gigantic jump with this giant digital leverage.

BDML can turn *Data* into *value*; Transform *information* into *intelligence;* Change *patterns* into *profit;* Convert *relationships* into *resources.* Companies and organizations can make *Faster* (real-time or near real-time), *Frequent*, and *Fact-based* decisions. In an ever-evolving market, 4IR with a set of technologies can stimulate innovations and rapid responses. Knowledge workers can proactively take action before an unfriendly event occurs (Wang, 2008).

Having been penetrated and integrated into almost every aspect of our work and life, as well as our society itself, AI and related cutting-edge technologies will enhance human capacities, improve efficiencies, and optimize people's lives. AI would not replace human intelligence, rather than amplify it. As *AI evolves* and *humans* adapt, AI and humans go forward together in the long run because AI and people both bring different capabilities to society.

According to Klaus Schwab, the World Economic Forum Founder and Executive Chairman, 4IR intellectualizes precipitous change to industrial and societal prototypes and processes in the 21st century due to increasing interconnectivity and smart automation and finally blurs the lines among the physical, digital, and biological worlds. Part of the 4IR is the manner in which all types of machines and devices interact, correspond, and cooperate with each other. Even though there will be obvious job losses due to the replacement of tasks that humans have conducted for years by autonomous machines and/or software. On the contrary, there could be new business opportunities and plenty of new jobs for controlling "the new electricity" (Philbeck & Davis, 2018; Moll, 2022).

There are 207 qualified full chapters among 271 accepted proposals. Finally, the 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 45 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, Iraq, Japan, Lebanon, Macau, Malaysia, 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. Let's take a peek at a few of them:

Jaydip Sen has published around 300 articles in reputed international journals and referred conference proceedings (IEEE Xplore, ACM Digital Library, Springer LNCS, etc.), and 18 book chapters in books published by internationally renowned publishing houses. He is a Senior Member of ACM, USA a Member of IEEE, USA. He has been listed among the top 2% scientists in the globe as per studies conducted by Stanford University for the last consecutive three years 2019 - 2021. In his contributed chapter Prof. Sen and his co-author, Dutta have evaluated the performance of two risk-based portfolio design algorithms.

Leung - who has authored more than 300 refereed publications on the topics of data science, ML, BDM and analytics, and visual analytics (including those in ACM TODS, IEEE ICDE, and IEEE ICDM) - presents two encyclopedia articles. One of them presents up-to-date definitions in BDM and analytics in the high-performance computing environment and focuses on mining frequent patterns with the MapReduce programming model. Another one provides the latest comprehensive coverage on key concepts and applications for BD visualization; it focuses on visualizing BD, frequent patterns, and association rules.

Lorenzo Magnani is Editor-in-Chief of the Series Sapere, Springer. Thanks to his logico-epistemological and cognitive studies on the problem of abductive cognition (that regards all kinds of reasoning to hypotheses) explained in this chapter both virtues and limitations of some DL applications, taking advantage of the analysis of the famous AlphaGo/AlphaZero program and the concepts of locked and unlocked strategies. Furthermore, he is the author of many important articles and books on epistemology, logic, cognitive science, and the relationships between ethics, technology, and violence.

The chapter 'AI is transforming insurance with five emerging business models' is the culmination of three years of research into how AI is disrupting insurance. Zarifis has recently won a 'best paper award' at a leading conference and Cheng has recently been published in MIS Quarterly for related work. AI is disrupting many distinct parts of our life, but insurance is particularly interesting as some issues like risk and privacy concerns are more important. After several case studies, this chapter identifies that there are five emerging models in insurance that are optimal for AI.

In "Artificial Intelligence, Consumers, and the Experience Economy," Chang and Mukherjee's excellent synthesis of AI and consumers in the modern economy provides a much-needed knowledge base for stakeholders tasked to deploy AI. In "Using Machine Learning Methods to Extract Behavioral Insights from Consumer Data," they present a comprehensive discussion of new data sources and state-of-the-art techniques for researchers and practitioners in computational social science. The chapters are built on their projects supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 Grant No. MOE2019-T2-1-183 and Grant No. MOE2018-T2-1-181, respectively.

Based on many years of application development by CY Pang and S. Pang's cognitive data analysis of many industrial projects, this chapter proposes a programming paradigm specific to BD processing. Pang was the lead architect of a $1.6 billion enterprise software project and was awarded a special architectural design trophy. He has received awards of $20,000 and $5,000 for outstanding innovation from a company he previously worked for. By the way, CY Pang was awarded a Prestige Scholarship from Peter House, Cambridge to complete his Ph.D. at the University of Cambridge, UK.

Vitor provides an excellent overview of multidimensional search methods for optimization and the potential these methods have to solve optimization problems more quickly. With almost ten years of industry experience, Vitor is an expert in optimization methods and the modeling of complex systems using operations research and data analytics techniques. He is also a recipient of the Nebraska EPSCoR FIRST Award, supported by the National Science Foundation to advance the research of early-career tenure-track faculty.

Lee's chapter on evidence-based data-driven pain management bears multi-facet importance. Nearly 40 million anesthetics are administered each year in the United States. And over 10.7% of Americans use prescription pain medication on a regular basis. The findings highlight the optimal safe dose and delivery mechanism to achieve the best outcome. The study showcases the persistence of overprescription of opioid-type drugs, as it finds that the use of fentanyl has little effect on the outcome and should be avoided.

Auditors must evaluate the volatility and uncertainty of the client company at the initial stage of the audit contract because it directly influences the audit risk. Takada contributes to auditing research and accounting education for 40 years. He has been awarded for his research and contributions to his excellent papers and accounting education by the *Chinese Auditing Association* and by the *Japanese Auditing Association*.

Nguyen and Quinn propose an optimal approach to tackle the well-known issue of the imbalance in bankruptcy prediction. Their approach has been evaluated through a rigorous computation including the most popular current methods in the literature. They have also made other main contributions in the area of imbalanced classification by winning the 2020 Literati Awards for Outstanding Author Contribution.

Rodríguez is the Bioethics of Displacement pioneer, a field that merges futurism, belongingness, and life. He has also published analytic papers and fieldwork on crises and big social changes such as pandemics, Anthropocene, AI takeover, cyborgs, digital securitization and terrorist attacks. As a chair, the author leads the research on the first decolonized corruption index. Torres shares his more than 15

years of wealth of experience in Predictive Maintenance management as a speaker at global summits such as Scalable and PMM Tech Dates. The author leads the first non-taxonomic error mode proponent of AI implementation.

Kurpicz-Briki, Glauser, and Schmid are using unique API technologies to measure the impact of online search behavior using several different online channels. Their method allows the identification of the specific channels, where keywords have been searched, and a restriction of regions, using the domains. Such technologies provide a major benefit for different application domains, including public health. In times such as a pandemic crisis, it is highly relevant for different stakeholders to identify the impact of their communications on the user community as well as the well-being of the population. Using the method proposed by the authors, this can be done while fully respecting the privacy of the users.

Sensors sense the environment and process large sets of data. Monitoring the data to detect malicious content is one of the biggest challenges. The previous work used mean variation to ease the surveillance of information. Ambika's proposal minimizes the effort by classifying the streamed data into three subsets. It uses the k-nearest neighbor procedure to accomplish the same. The work conserves 10.77% of energy and tracks 27.58% of more packets. Map-reduce methodology manages large amounts of data to a certain extent. Ambika's other proposal aims to increase processing speed by 29.6% using a hashing methodology.

In today's world, text-based sentiment analysis brings the attention of all. By looking at the people requirement, Tripathy and Sharaff propose a hybridized Genetic Algorithm (GA)-based feature selection method to achieve a better model performance. In the current study, they have customized the GA by using the SVM to evaluate the fitness value of the solutions. The proposed idea is essential as the technique reduces the computational cost by reducing sufficient features without affecting the performance. The proposed model can be implemented in any field to filter out the sentiment from the user's review.

Alberg and Hadad present the novel Interval Gradient Prediction Tree ML Algorithm that can process incoming mean-variance aggregated multivariate temporal data and make stable interval predictions of a target numerical variable. Empirical evaluations of multi-sensor aircraft datasets have demonstrated that this algorithm provides better readability and similar performance compared to other ML regression tree algorithms.

The environmental, societal, and cultural imperatives press for innovative, prompt, and practical solutions for grave humanitarian problems we face in the 21st century. The climate crisis is felt everywhere; natural disasters are rampant. Can technology provide reasonable means to humanitarian supply chains? What potential uses can AI offer in establishing sustainable humanitarian logistics (SHL)? Ülkü, an award-winning professor and the director of CRSSCA-Centre for Research in Sustainable Supply Chain Analytics, and his research associate Oguntola of Dalhousie University - Canada review the latest research on the applications of AI technology on SHL.

Aguiar-Pérez, the leading author of this chapter, provides the audience an insight into what ML is and its relation with AI or DL. He has an extended experience in the field of ML, DL, BD, and IoT in various sectors (automotive, smart roads, agriculture, livestock, heritage, etc.), including collaboration with companies, EU-funded research projects, publications, and postgraduate teaching experience. The rest of the authors work with him in the Data Engineering Research Unit of the University of Valladolid.

Bagui, a highly accomplished author of several books on databases and Oracle, presents a very timely chapter on the improvements made in Oracle 19c's multitenant container architecture and shows how these improvements aid in the management of Big Data from the perspective of application development. The added functionality that comes with the integration of Big Data platforms, alongside the flexibility

and improvement that comes with a container and pluggable databases, has allowed Oracle to be in the forefront in the handling of Big Data.

As an internationally renowned interdisciplinary information and data professional, Koltay's chapter on Research Data Management (RDM) is of interest not only for both professionals of DS and ML but is related to any research activity. He is also a widely published author in these fields. In 2021, his contribution to IGI Global books included an entry on information overload. His book, titled Research Data Management and Data Literacy (Chandos, 2021) contains a more detailed explanation of the subjects, contained in this chapter.

Zhao is a DS professional with experience in industry, teaching, and research. He is a leading BD expert in the IR BD & AI Lab in New Jersey, USA. He provides multiple chapters to the book by covering a broad range of BD applications in vast perspectives of urgent demands in DS research objectives, such as DSS, DL, computer vision, BD architecture designs, and applied BD analytics in Covid-19 research. As such, he did excellent work in those chapters and made significant contributions to the book.

Based on their discovery of action rules and meta-actions from client datasets, Duan and Ras propose a strategy for improving the number of promoters and decreasing the number of detractors among customers. Moreover, the improved/enhanced action rules can be utilized in developing actionable strategies for decision makers to reduce customer churn, which will contribute to the overall customer churn study in the business field. The authors target the domain represented by many clients, each one involved with customers in the same type of business. Clients are heavy equipment repair shops, and customers are owners of such equipment.

The A2E Process Model for Data Analytics is simple without being simplistic and comprehensive without being complicated. It balances technology with humanity and theories with practices. This model reflects Jay Wang's decades-long multi-disciplinary training and experience in STEM, Behavioral Science, and Management Science. While existing process models such as CRISP-DM, SEMMA, and KDD were developed for technical professionals with limitations and low adoption rates, the A2E Model is more approachable to subject matter experts, business analysts, and social scientists. The A2E Model will elevate the analytics profession by fostering interdisciplinary collaborations of all stakeholders and increasing the effectiveness and impacts of analytics efforts.

Turuk explores Audio and video-based Emotion Recognition using the Backpropagation Algorithm, which is the backbone of ML and DL architectures. This chapter analyses everyday human emotions such as Happy, Sad, Neutral, and Angry using audio-visual cues. The audio features such as Energy & MFCC and video features using the Gabor filter are extracted. Mutual information is computed using video features. The readers will benefit and motivated to conduct further research in this domain. The application may be extended to a lie detector using Emotions.

Stojanović and Marković-Petrović focus on continuous cyber security risk assessment in Industrial Internet of Things (IIoT) networks, and particularly on possibilities of DL approaches to achieve the goal. The authors successfully complement their previous work regarding the cyber security of industrial control systems. They concisely review the theoretical background and provide an excellent framework for the continuous risk assessment process in the IIoT environment. DL can be integrated into edge-computing-based systems and used for feature extraction and risk classification from massive raw data. The chapter ends with a list of proposals for further studies.

Climate change is a very important issue and each person on our planet must have a culture of keeping it clean. Pollution increased yearly due to the increased consumption of fossil fuels. Alsultanny has many research papers in climate change and renewable energy. He led a UNDP team for writing reports

on energy consumption in Bahrain. Alsultanny did an innovative method in his chapter, by utilizing the pollution gases data, these data currently are BD, because they are registered yearly in every minute, and from many monitoring pollutions stations.

Deliyska and Ivanova conducted timely research and practical work representing an important contribution to data modeling in sustainable science. Applying ontological engineering and a coevolutionary approach, a unique metamodel of sustainable development is created containing structured knowledge and mutual links between environmental, social, and economic dimensions in this interdisciplinary area. Specialists in different fields can use the proposed metamodel as a tool for terminology clarification, knowledge extraction, and interchange and for the structuring of ML models of sustainable development processes.

Hedayati and Schniederjans provide a broad spectrum of issues that come into play when using digital technologies to benefit healthcare. This is even more important where the pandemic has forced healthcare models to rapidly adjust towards compliance with local, regional, and national policy. The dissemination and creation of knowledge become paramount when considering the benefits and drawbacks of the rapid changes in technology applications worldwide. The authors consider several insights from the American Hospital Association Compliance to provide some questions researchers and practitioners may consider when addressing knowledge management via digital technology implementation in healthcare settings.

Pratihar and Kundu apply the theory of fuzzy logic to develop a classification and authentication system for beverages. It emphasizes the versatility of fuzzy logic to deal with the higher dimensional and highly non-linear sensor data obtained from e-tongue for different beverage samples. Commonly used mapping techniques (for dimension reduction of a data set) and clustering techniques (for classification) were also briefly discussed. This study provides a perspective on developing a fuzzy logic-based classifier/authenticator system in the future for beverages, foods, and others and their quality control and monitoring.

Drake discusses the use of IoT technology to improve SCM. As firms look to improve their supply chain resilience in response to the COVID-19 pandemic and other disruptions, IoT data increases visibility, traceability, and can help firms to mitigate risks through added agility and responsiveness. The improved decision-making made possible by IoT data creates a competitive advantage in the market.

Today, high-dimensional data (multi-omics data) are widely used. The high dimensionality of the data creates problems (time, cost, diagnosis, and treatment) in studies. Ipekten et al. introduce the existing solutions to these problems and commonly used methods. Also, the authors present the advantages of the methods over each other and enlighten the researchers that using suitable methods in terms of performance can increase the reliability and accuracy of the studies. Finally, the authors advise on what can be done in the future.

Learning analytics (LA), a promising field of study that started more than a decade ago but has blossomed in recent years, addresses the challenges of LA specifically in education, integrating it as a fundamental element of the Smart Classroom. Ifenthaler and Siemens among others discuss the primary features, the benefits, and some experiences. In addition, the team of authors of the chapter has contributed more than twelve publications on this topic in the last 3 years in leading journals and publishers.

Current advances in AI and ML in particular have raised several concerns regarding the trustworthiness and explainability of deployed AI systems. Knowledge-Based approaches based on symbolic representations and reasoning mechanisms can be used to deploy AI systems that are explainable and compliant with corresponding ethical and legal guidelines, thus complementing purely data-driven approaches.

Batsakis and Matsatsinis, both having vast theoretical backgrounds and experience in this research area, offer an overview of knowledge-based AI methods for the interested AI practitioner.

Noteboom and Zeng provide a comprehensive review of applications of AI and ML and data analytics techniques in clinical decision support systems (CDSSs) and make contributions including, 1) the current status of data-driven CDSSs, 2) identification and quantification of the extent to which theories and frameworks have guided the research, 3) understanding the synergy between AI/ML algorithms and modes of data analytics, 4) directions for advancing data-driven CDSSs to realize their potential in healthcare.

Fisogni investigates the emotional environment which is grounded in any human/machine interaction. Through the lenses of metaphysics and system thinking the author sketches a highly valuable insight, for sure an unprecedented challenge for DSs. In fact, only a philosophical foundation of the big issues of this realm can bring about a change in the quality of understanding an increasingly melted environment humans/machines in the Onlife era.

In "Hedonic Hunger and Obesity", Demirok and Uysal touch upon a remarkable topic and explain ways of identification for people with hedonic nutrition and the conditions that are effective in the states that trigger hunger state in humans. In addition, in this text, the authors ensample hormones that suppress and trigger hunger.

Yen and her coauthors contributed a chapter on how ML creates the virtual singer industry. Virtual singers have great market potential and even advantages over their human counterparts. Despite the bright future of virtual singers, the chapter has discussed difficulties virtual singers face, especially their copyright protection by legislation. Literature on the technical aspects of virtual singers is also reviewed, and a list of additional readings is provided for readers interested in the ML algorithms behind virtual singers.

Rastogi is working on Biofeedback therapy and its effect on Diabetes diseases, a currently very active healthcare domain. He brings back the glory of Indian Ancient Vedic Sciences of Jap, Pranayama, Healing techniques, and the effect of Yajna and Mantra science on Diseases and pollution control. Also, He has developed some interesting mathematical models with algorithms on Swarm Intelligence approaches like PSO, ACO BCO, etc. for better human life via Spiritual Index and higher consciousness.

Isikhan presents a comparison of a new proposal for the modeling of Ceiling and Floor Effect dependent variables and classical methods. It has been noticed that there are very few publications evaluating the regression modeling of ceiling and floor effect observations in recent years. The modeling method with regression-based imputation, which clinicians can use as an alternative to classical models for ceiling and floor effective observations, is explained in detail. The performances of the newly proposed imputation-based regression and other classical methods were validated based on both real clinical data, synthetic data, as well as a 500 replicated cross-validation method.

Drignei has extensive experience with time series modeling and analysis. Prior to this work, he addressed statistical modeling aspects of space-time data, such as temperatures recorded over space and time. His research has been published in leading statistics journals. The current work deals with seasonal times series recorded at a large number of time points. Such data sets will become more common in the future, in areas such as business, industry, and science. Therefore, this chapter is timely and important because it sheds new light on modeling aspects of this type of data sets.

Data visualization plays a key role in the decision-making process. Visualization allows for data to be consumable. If data is not consumable, there is a tendency to ignore the facts and rely more on biases. Researchers have found that cognitive biases do exist within data visualizations and can affect decision-making abilities. Anderson and Hardin provide background on cognitive biases related to data visualizations, with a particular interest in visual analytics in BD environments. A review of recent

studies related to mitigating cognitive biases is presented. Recommendations for mitigating biases in visualizations are provided to practitioners.

Puzzanghera explores the impact of AI on administrative law. He combines IT systems with administrative activity and researches the processors that prepare content and the implications that arise. He analyzes the European Commission's proposal in regard to the legislation of AI in Europe and the importance of safeguarding human rights in the introduction of AI in administrative activity.

How ML impacts the catering industry? Liu et al. provide a comprehensive vision to readers with real-life examples and academic research. Researchers at business schools may have their attention drawn to the impact of ML on operations, management, and marketing, while scholars with solid ML backgrounds may become aware of industry issues, identify new research questions, and link their expertise to practical problems through reading the chapter.

Di Wang's research interests include 4D printing technology, robot control, remanufactured industry, and energy schedule in the smart city. Combinatorial optimization is a widely applied field at the forefront of combinatorics and theoretical computer science. With BD challenges, deep reinforcement learning opens new doors to solve complex combinatorial optimization problems with overwhelming advantages over traditional methods.

Firmansyah and Harsanto focus on exploring BD and Islamic finance. The utilization of BD in Islamic financial institutions (IFIs) has been perceived as a source of competitive advantage in today's era. Many IFIs have been more dependent on BD technologies than ever before in order to keep up with the changing customers' demands, lifestyles, and preferences.

With his experience of working in both industry and academic research, Indraneel highlights progress made in integrating AI with industry and helps bridge the reality and challenges faced while summarizing the state of Industry 4.0. The author engages audiences from different sectors without overburdening the reader with incoherent technical details. A practitioner in the fields of DS and cybersecurity, the author brings experience interacting with clients and customers from different fields, including manufacturing, legal, and product developers.

Yang, Wu, & Forrest examine the textual aspects of consumer reviews. As a critical source of information for online shoppers, researchers have spent considerable time examining the potential impact of consumer reviews on purchasing behavior. The authors contribute to the existing body of knowledge by proposing a conceptual framework for capturing the internal relationships between major textual features discovered in prior research.

Kara and Gonce Koçken are researchers studying mathematical programming problems in fuzzy environments. In the study, a novel fuzzy solution approach to multi-objective solid transportation problems is developed by using different membership functions, which can help the studies in transportation systems.

Millham demonstrates the various spheres of the emerging 4IR and how they interrelate with the application, opportunities, expectations, and challenges of a smart city. Because many of these smart city applications are very complex and interact with each other using various technologies, several nature-inspired algorithms are introduced as a way to provide intelligent and coordinated management of these entities.

The development of novel measurement and detection techniques is a rapidly growing area, where the generation of vast amounts of information requires novel methods for analysis. Murrieta-Rico explores a new direction of his research by combing the know-how for generating a big dataset from a digital frequency measurement, with the application of the principal component analysis (PCA). As a result, a

powerful methodology for data analysis is presented. In addition, these results can be used for extending the capabilities of ML systems based on sensors.

Coimbra, Chimenti, and Nogueira contribute to the debate related to human-machine interaction in social media. The work helped to understand the mechanisms and motivators of this relationship. In addition, the article presented a historical evolution of the debate on the interaction between machines and men in decision-making, distributing the result of the literature review in three historical cycles. The research was carried out through a survey of YouTube users to understand the interaction mechanism along with its motivators.

As a transformational general-purpose technology, AI is impacting marketing as a function, and marketing managers' activities, capabilities, and performance. Oberoi emphasizes how the job of a marketing manager will be evolving into understanding which kind of AI can and should be applied to which kind of marketing actions for better performance. Marketing managers will have to go through a learning curve and acquire new skills.

Singh and Dev have discussed the concepts of data warehouse and OLAP technology to deal with real-life applications efficiently. The topic is useful in the modern digital era as businesses are dealing with data from heterogeneous sources. The chapter presents the case study of the tourism industry as it deals with multidimensional data like tourist, hospitality, and tourist products. This chapter will be helpful in understanding how to generate multi-dimensional reports that will show the information according to the needs of policymakers.

Ramos has made many contributions to the potential of Business Intelligence tools, combined with DM algorithms methods to produce insights about the tourism business, highlighting an aspect of the investment potential of tourism organizations in this type of system, from those related to accommodation, management of tourist destinations, to tourist transport, restaurants, among other businesses complementary to the tourist activity, with a view to innovation and increasing financial performance, which includes examples ranging from the application of OLAP techniques to the application of ML methods.

Balsam depicts the meaning and role of metamodels in defining the abstract syntax of the language by which developers communicate, design, and implement systems including the selection of the design, implementation methods, and techniques for increasingly complex systems to satisfy customers' needs, particularly if the system has to be delivered in a considerably fleeting time. The author highlights different aspects of meta-models standards, categories, the process of creating the metamodel, and challenges in the research of metamodeling.

Dharmapala contributes a novel method to the field of research in 'Classification of employee categories in allocating a reward, with input features from survey responses.' In the past, researchers conducted qualitative and quantitative analyses on this subject as it is an important topic to any organization that strives to boost the morale of its employees. The author opened a new direction in future research on the subject by using ML algorithms, and the results obtained were promising.

Mudrakola identifies the gap and future scope for Breast cancer applications like the impact of chemical therapy, prognosis analysis among various treatment types and stages, etc. From basic to the latest trends, the author's extensive literature survey will direct the root to aspects needed to analyze work on medical applications specific to Breast cancer.

Rani et al. highlight the venues of user-generated content (UGC) in Industry 4.0. This chapter's contribution is highly interesting for any digital content creator and non-paid professionals. The importance of UGC on consumer behavior in the era of Industry 4.0 will be explained, allowing stakeholders to assess their efficacy in Internet communication and enhancing the digital process required for modern

marketing. The chapter aims to link existing ideas and provide a holistic picture of UGC by concentrating on future research.

Ibrahim et al. seek to provide an understanding of the relationship between member support exchange behavior and self-disclosure intention in online health support communities using a data-driven literature review. Seeking or providing support in online communities may be useful but having to disclose personal information publicly online is a critical privacy risk – intention counts.

Rusko introduces the main perspectives of industrial revolutions. He found interesting backgrounding details for the chapter about the disruptions of the industrial revolutions. Kosonen updates the paper with the effects of Covid-19 and contemporary digitizing development.

I would like to highlight a number of authors who have received special stunning honors: Eva K Lee has published over 220 research articles, and fifty government and state reports, and has received patents on innovative medical systems and devices. She is frequently tapped by a variety of health and security policymakers in Washington for her expertise in personalized medicine, chronic diseases, healthcare quality, modeling and decision support, vaccine research and national security, pandemic, and medical preparedness. Lee has received multiple prestigious analytics and practice excellence awards including INFORMS Franz Edelman award, Daniel H Wagner prize, and the Caterpillar and Innovative Applications in Analytics Award for novel cancer therapeutics, bioterrorism emergency response, and mass casualty mitigation, personalized disease management, ML for best practice discovery, transforming clinical workflow and patient care, vaccine immunity prediction, and reducing hospital-acquired infections. She is an INFORMS Fellow. She is also inducted into the American Institute for Medical and Biological Engineering (AIMBE) College of Fellows, the first IE/OR engineer to be nominated and elected for this honor. Her work has been funded by CDC, HHS, NIH, NSF, and DTRA. Lee was an NSF CAREER Young Investigator and Whitaker Foundation Young Investigator recipient.

Petry and Yager are both internationally known for their research in computational intelligence, in the area of fuzzy set theory and applications, and are both IEEE Fellows and have received prestigious awards from the IEEE. They have collaborated here as it represents extensions of their previous research on this topic. Hierarchical concept generalization is one important approach to dealing with the complex issues involving BD. This chapter provides insights on how to extend hierarchical generalization to data with interval and intuitionistic forms of uncertainty.

The globalization of the software development industry continues to experience significant growth. The increasing trend of globalization brings new challenges and increases the scope of the core functions of the software development process. Pal introduces a distributed software development knowledge management architecture. Kamalendu has published research articles in the software development community in the ACM SIGMIS Database, Expert Systems with Applications, DSSs, and conferences. Kamalendu was awarded the best research paper on data analytic work at a recent international conference. He is a member of the British Computer Society, the IET, and the IEEE Computer Society.

Badia's research has been funded by the National Science Foundation (including a prestigious CAREER Award) and has resulted in over 50 publications in scientific journals and conferences. His chapter demonstrates how to use SQL in order to prepare data that resides in database tables for analysis. The reader is guided through steps for Exploratory Data Analysis (EDA), data cleaning (including dealing with missing data, outliers, and duplicates), and other tasks that are an integral part of the Data Scientist day-to-day. The references provide a guide for further study.

Srinivasan explains the three components of graph analytics and provides illustrative examples as well as code for implementation. His chapter is one of the few primers of graph DS/analytics that covers a variety of topics in the discipline. The author does active research in graph analytics methods and applications in healthcare, ML explainability, and DL and regularly publishes in top journals and conferences in information systems, healthcare, and computer science. He received best paper awards in INFORMS Workshop on Data Science (2021) and the 6th International Conference on Digital Health (2016), respectively.

Knowledge explosion pushes BDML, a multidisciplinary subject, to ever-expanding regions. Inclusion, omission, emphasis, evolution, and even revolution are part of our professional life. In spite of our efforts to be careful, should you find any ambiguities of perceived inaccuracies, please contact me at prof.johnwang@gmail.com.

John Wang
Montclair State University, USA

REFERENCES

Moll, I. (2022). The Fourth Industrial Revolution: A new ideology. *tripleC: Communication, Capitalism & Critique*, *20*(1), 45–61.

Philbeck, T., & Davis, N. (2018). The Fourth Industrial Revolution: Shaping a new era. *Journal of International Affairs*, *72*(1), 17–22.

Roy, D., Srivastava, R., Jat, M., & Karaca, M. S. (2022). A complete overview of analytics techniques: Descriptive, predictive, and prescriptive. *Decision Intelligence Analytics and the Implementation of Strategic Business Management*, 15-30.

Wang, J. (Ed.). (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (Vols. 1–6). IGI Global. doi:10.4018/978-1-59904-951-9

Wang, J. (Ed.). (2022). *Encyclopedia of Data Science and Machine Learning*. IGI Global. https://www.igi-global.com/book/encyclopedia-data-science-machine-learning/276507

Table of Contents by Category

Section: Big Data Applications

Section: Big Data as a Service

Section: Big Data Systems and Tools

Section: Business Intelligence

Volume II

Section: Causal Analysis

Section: Chaos Control, Modeling, and Engineering

Section: Cloud Infrastructure

Section: Cognitive Science

Section: Computational Intelligence

Section: Computational Statistics

Section: Computer Vision

Section: Customer Analytics

Section: Data Processing, Data Pipeline, and Data Engineering

Volume III

Section: Data Visualization and Visual Mining

Section: Decision, Support System

Section: Deep Neural Network (DNN) of Deep Learning

Section: E-Learning Technologies and Tools

Section: Emerging Technologies, Applications, and Related Issues

Section: Ensemble Learning

Section: Gradient-Boosting Decision Trees

Section: Graph Learning

Section: High-Throughput Data Analysis

Section: Industry 4.0

Section: Information Extraction

Section: Internet of Things

Section: Malware Analysis

Section: Management Analytics

Section: Marketing Analytics

Section: Mathematical Optimization

Section: Meta-Analysis and Metamodeling

Section: Multivariate Analysis

Section: Natural Language Processing (NLP)

Section: Nature-Inspired Algorithms

Volume V

Section: Network Modeling and Theory

Section: Object Detection

Section: Performance Metrics

Section: Predictive Analytics

Section: Pricing Analytics

Section: Qualitative Research

Section: Recommender Systems

Section: Reinforcement Learning

Section: Simulation and Modeling

Section: Smart City

Section: Social Media Analytics

Section: Supply Chain Analytics and Management

Section: Symbolic Learning

Section: Time Series Analysis

Section: Transfer Learning

Section: Transport Analytics

Section: Unsupervised and Supervised Learning

The Role of Metamodeling in Systems Development

Balsam A. J. Mustafa
https://orcid.org/0000-0002-2973-7124
Al Hadi University College, Iraq

Mazlina Abdul Majid
University Malaysia Pahang, Malaysia

INTRODUCTION

What is Meta-Model? The word *"Meta"* literally means "after", "beyond" in Greek or more "comprehensive" (*Merriam-Webster's*, 1993). In computer science, it is widely used in different meanings. In Database Management Systems, metadata means (data about data) which may represent data dictionaries, repositories, etc., and models represent data like the ER model (Entity-Relationship). In other words, a meta-model is a model of the data model. In Programming Languages, a meta interpreter is an interpreter of a (program) interpreter (Smith, 1984). Meta-modeling is a method for defining the abstract syntax of a language, both modeling or programming language. It makes the development of a language simpler allowing the designers to directly map the classes identified in domain analysis to classes in the meta-model (Kleppe, 2008). The meta-model expresses what models include such as concepts and relationships between them and may be the rules of how these concepts can be interrelated. Hence, a metamodel can be treated just like any conceptual model of information systems. The only specialty is that the artifact of meta-modeling is a model (Jeusfeld, 2009), i.e. a model is an instance of a metamodeling, for example, any UML class diagram can be seen as an instance of the UML metamodel that should be well formed with respect to it (Osis & Donins, 2017).

At the beginning of the development of any system, the representation of a system view takes place once it is represented by a model. Kuhne (2006). defined a model as an abstraction of a real or language-based system allowing predictions or inferences to be made by developers. Kuhne (2006). explained that any model is built according to a specific meta-model which consists of a collection of functional or structural elements and rules to allow modeling the system view. The developer then can explain his ideas and discuss the conceptual view of the system with other stakeholders and can be further refined based on feedback from others. For any modeling method, its accuracy depends on the meta-model which semantically supports the features and behavior of the system that the method is used to model, i.e. the metamodel identifies the semantics of the system representation at the model level. Basha et al. (2012). pointed out that metamodeling is important because it provides a means for the machine to read, write, and understand models that were created and interpreted only by people. From this perspective, meta-modeling plays a key role in automating model based system development (MBSD). With models understandable to computers, tools can be built for model creation and code generation.

DOI: 10.4018/978-1-7998-9220-5.ch145

Harel and Rumpe (2004). explained that machines use machine-readable languages for communication just like people use natural languages to communicate between them. Both kinds of languages whether they are natural, artiðcial, or programming languages contain a large variety of meaningful elements. Therefore, a language consists of a syntactic notation (syntax), which is a set of legal elements, together with the meaning of those elements, which is expressed by relating the syntax to a semantic domain. A metamodel is a model of a language that developers use to design and implement a system and its structure consists of the essential elements of the language such as the language concepts, its graphical syntax and its semantics, i.e. what the models and programs written in the language mean and how they behave (Atkinson & Kuhne, 2003).

To improve system development practice, it is important to understand how languages can be managed to respond to the developers' demands. Language driven development is elaborated in this chapter. The right language improves the productivity of developers by increasing the value of primary software artifacts in terms of how much functionality they deliver at the development stages, and by reducing the rate at which primary software artifacts become obsolete (Atkinson & Kuhne, 2003).

In this chapter, we discuss first the role and importance of a language used in system development with its features, and introduces the steps of the meta-modeling process. In the sequel, we review the meta-modeling standards and the relationship between meta-modeling and model-driven architecture (MDA). The chapter also discusses the different categories of meta-models and finally presents the areas in meta-modeling where there is a demand for more research suggesting some future work.

BACKGROUND

The architecture of a system involves what elements make up the system and how they work together to provide the functionality of the system. The Model Driven Architecture (MDA) is an approach for software systems development initiated by the object management group (OMG) in 2001 (OMG, 2001). Unlike the other standards of the OMG, the MDA offers to use models instead of the traditional source code. It defines a specification that separates the system functionality from implementation that uses a specific technology platform. The architecture of a system is a specification of the parts and connectors of the system and the rules for the interactions of the parts using the connectors (Shaw, & Garlan, 1996). The standard of MDA released by OMG contains a set of guidelines for structuring the specification as models. MDA focused on creating and modeling the software products (Yousaf et al, 2019). The model here is an abstraction or a representation of a certain aspect of the system or a domain of the real-world that the system is designed for. The model aims to simplify the complexity of the system, focuses on the abstraction away from code to form a graphical model, which enables developers to understand, communicate, design, and implement the systems or adapt existing models. This makes the development of an application easier for those without prior coding knowledge. Different standard notations are used in modeling, e.g. the unified modeling language (UML).

MDA hence relies on models to be the main artifact in the development process to raise the level of abstraction to manage the complexity and change of the development process. This includes all types of models defined in the OMG standard which are the platform-independent model (PIM) and the platform-dependent model (PDM) to cover all system aspects in the development lifecycle. Models are created by a language, and the OMG' Meta-Object Facility (MOF) is introduced and clearly stated as the language in which all the languages for MDA are written (MOF, 2002). Modeling languages are used to define models, thus, their syntax and semantics (meaning) must be precisely defined. Hence, the process of

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z