THE WRITING CENTER

Data Analytics Engineering

Writing Center Guide

Data Analytics Engineering is a multidisciplinary field that applies relevant algorithms and techniques to other application domains, such as business, finance, marketing, physics, biology, geography, etc. It aims to extract useful and valuable information from collected datasets, make credible prediction for future, and help managers do data-driven decision-making. In particular, these techniques mainly involve domains of statistics, operations research, data mining, machine learning, deep learning, database, big data, visualization, etc. This brochure will further introduce research methods and writing standards in the data analytics engineering domain.

- **Values**
- **Current Pressing Issues**
- **Research Questions**
- **To Do Research**
- **Writing**

Values

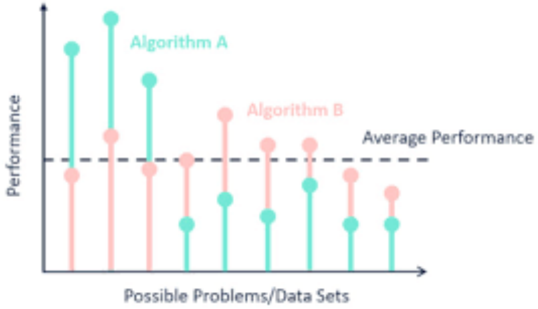
Data analysis not only requires technical skills but ideas. Some values can guide us to pursue the best results in the right way.

Occam's Razor Principle

Occam's Razor Principle means the simplest explanation is the best explanation [1]. For example, if there are two models, a simple one and a complex one, that can address the same problem, the wise choice is the simple one [1]. This is because, as for the data science domain, it is easier for a complex model to produce the overfitting than the simple one [1]. What is overfitting? Overfitting, one of the most important evaluation strategies, means that the model has an excellent performance on the training data set, but a high error rate on the independent testing dataset [1]. Hence, Skiena encourages us to start with the easy methods and try the advanced ones if the potential improvements in accuracy really justify it [1].

No Free Lunch Theorem

No Free Lunch Theorem emphasizes, "There does not exist a single machine learning algorithm better than all the others on all problems [2]." In other words, certainly no single machine learning method dominates all the others [2]. Therefore, we cannot look forward to an algorithm which can address all problems or always performs better than other algorithms. What we need to do is to select the most appropriate model to obtain the most optimal solution and performance according to the specific situation, predefined conditions, and certain purpose. Moreover, it depends on our comprehensive understandings of the problem and each algorithm.



Current Pressing Issues

As for the current pressing issues, this guide summarizes them as two aspects which are technical level and application level.

Technical Level

Currently, the appearance of big data brings both opportunities and challenges to data science. Undeniably, there exists enormous valuable information in big data that remains to be explored. However, the three properties of big data [3], which are the large scale of data volume, the variety of data types, and the fast velocity of updating, put forward higher demands on the form of data management and access [4], and efficient algorithms to work on the data [5]. For example, up to now, there are about 4.4 billion Internet users in the world [6] who produce overwhelming data. On average, 500 million tweets are sent per day [7] which include data types of text, video, sound, etc. Moreover, there are 3.5 billion Google searches per day which present around 65% of all web searches worldwide [8]. Nowadays, small companies even handle exploded data including customer behavior, market characteristics, product inventories and more can all be tracked in real-time [4].

- As for **the form of data management and access**, proper management approaches can more easily scale the data operation with business and share valuable insights across the organization, from IT to analysts to executives [4]. However, the real situation of most companies is lacking an efficient organization among data collection, delivery, analyzation, and visualization, which further leads to unnecessary time loss in terms of business intelligence and data analytics [4]. More recently, the rise of cloud computing provides more efficient data management solutions [4]. In addition, the development of Data Warehouse and Data Lake offers more advanced data storage tools [4].
- **The efficient algorithms to work on big data** aim to reduce the time complexity and time cost of running [5]. However, studies on this aspect are still underway. Specifically, the basic methods mainly involve asymptotic complexity, hashing, and streaming models to optimize I/O performance in large data files [5]. In addition, parallel computing and distributed computing are two solutions that provide simultaneously computing with multiple machines [5]. For example, the cloud computing platforms, such as Amazon AWS, Google Cloud, and Microsoft Azure, provide paid service of simultaneously computing ability [5]. Another example is that Google's MapReduce paradigm, based on open-source implementations like Hadoop and Spark, is an accomplishment of distributed computing [5].

In addition to the above technical issues, data scientists still pursue more efficient and flexible data ingestion, cleaning, and transformation solutions [4]. Moreover, more advanced data analyzation methods and interactive visualization are also two important issues in the data science domain [9].

Current Pressing Issues

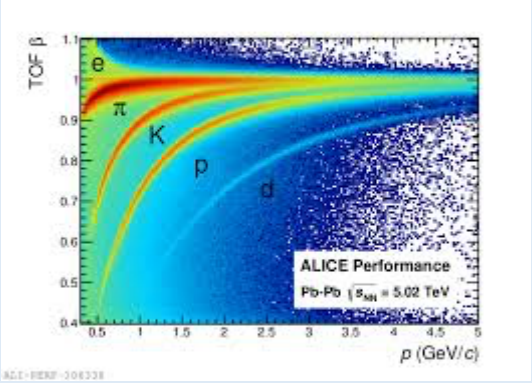
As for the current pressing issues, this guide summarizes them as two aspects which are technical level and application level.

Application Level

This guide just illustrates issues of three application domains, which are the finance domain, the particle physical domain, and the medical domain.

In **the finance domain**, risk management is one of the most important issues, including operational risk, credit risk, market risk, volatility risk, etc. [10]. For instance, towards both clients and companies, the credits classification and default prediction are ideal approaches to avoid credit risk [10]. Tsai and Wu use diversified multiple classifiers based on neural networks to make bankruptcy prediction and credit scoring [11].

In **the particle physical domain**, particle identification is one of the most important problems [12]. It can be regarded as a classification issue in the machine learning domain. For instance, Roe and Yang applied a particle identification technique based on decision trees to the MiniBooNE experiment, an experiment at Fermilab searching for neutrino oscillations [12].



In **the medical domain**, cancer is a disease that has a great impact on society in the United States and across the world [13]. Though there have been overwhelming existing real-time electronic health records of cancer, doctors cannot easily analyze them and apply them to the medical diagnosis [14]. The collaboration between IBM Watson supercomputer and Memorial Sloan-Kettering Cancer Center is trying to use the big data analyzation methods to guide the clinical decision [14]. If it finally works, physicians not only will stay current with the latest evidence guiding clinical but also will not be worried about the difficulty of the translation of a great deal of knowledge and sorting through that information [14].

Research Questions

First of all, all the techniques mentioned above serves the applications in other domains. Though the applications involve a great number of fields, we can mainly summarize all research questions as following types: ***statistical analyzation*** [15] [16], ***optimization problems*** [17], ***association rule*** [18], ***regression*** [19], ***clustering*** [20], ***classification*** [2] [20], and ***anomaly detection*** [21].

- **Statistical analyzations** mainly involve the understanding of data and statistical prediction, particularly the experiments that are hard to make in real life [15]. For instance, Hornung predicts the occupational mortality of retrospective exposure to ethylene oxide via a statistical prediction model [22].
- **Optimization problems** mainly include scheduling and time management, enterprise resource planning (ERP) and supply chain management (SCM), inventory management, network optimization and engineering, packet routing optimization, risk management, etc. [23].
- **Association rule** is a problem of investigating the relationship between variables in the dataset [18]. For instance, the association rule learning of behaviors of purchasing commodities of customers, such as bread and milk, can provide marketing managers with valuable guidance [24].
- **Regression problem** means making the value prediction via collected data [19]. For example, based on the GDP records of years before, we can predict future GDP by regression models [19].
- **Clustering** is a problem of grouping points by similarity [20]. It is more like an understanding of the data structure [20]. For example, based on the images which have no class labels, distinguishing some characteristic landcover types is an application of clustering [25].
- **Classification problem** means to predict the categories of data based on pre-collected data [2]. For instance, according to a great deal of collected data from numerous companies, we can build a model for bankruptcy prediction and credit scoring [11]. In this case, the categories are whether the company has the risk to go bankrupt or not [11].
- **Anomaly detection** is a problem of identifying rare items, events, or observations in the database [26]. For instance, efficiently distinguishing the outlier records of data is one of the most important approaches to fraud detection [27].

Having a deep understanding of the above problem types and distinguishing them well is a good beginning of a data analysis project. Sometimes, a specific issue may involve more than one problem type and we need to combine multiple methods to address it.

To Do Research

To do research in the data analytics domain, the processes usually involve asking an interesting question, getting the data, exploring the data, modeling, evaluating and analyzing, communicating and visualizing the results. ***This brochure will further discuss methods of the collection of data, modeling, and evaluation.***

The Collection of Data

Open Source Database

Here list some open databases that can be used in the research of the data science domain.

1) **Google's Dataset Search Engine**: https://toolbox.google.com/datasetsearch
2) **Registry of Open Data on Amazon**: https://registry.opendata.aws/
3) **Open Image Database (ImageNet):** http://www.image-net.org/
4) **World Bank Open Data:** https://datacatalog.worldbank.org/
5) **UCI Machine Learning Repository:** https://archive.ics.uci.edu/ml/index.php

In addition, Mason Digital Scholarship Center provides the comprehensive Data InfoGuides: https://dsc.gmu.edu/data/

The Database of Scholar Source

Core databases of scholar sources for data analytics engineering are listed on the website of InfoGuides on Mason Library [28]: https://infoguides.gmu.edu/DAEN/articles. For example:

1) **Association for Computing Machinery(ACM) Digital Library** [29]

Articles in this database involve the area of computing and information technologies [28]. It has multiple types of publications, such as journals, magazines, and conference proceedings [28].

2) **Computers and Applied Sciences Complete** [30]

This database covers subject areas include the computing and applied sciences disciplines, engineering, computer theory and systems, new technologies, etc. [28].

3) **IEEE Xplore** [31]

This database provides full-text access to all journals, transactions and conference proceedings published by IEEE, and IET [28].

4) **ScienceDirect** [32]

This database access journals and eBooks published by Elsevier and its partners. Covers most disciplines, including science, medicine, economics, and the social sciences [28].

5) **Springer LINK** [33]

This database concludes subjects in computer science, engineering, humanities, social sciences, law, business and economics, biomedical, and life sciences [28].

To Do Research

To do research in the data analytics domain, the processes usually involve asking an interesting question, getting the data, exploring the data, modeling, evaluating and analyzing, communicating and visualizing the results. ***This brochure will further discuss methods of the collection of data, modeling, and evaluation.***

The Collection of Data

Scholarly Journals

1) **Artificial Intelligence** [34]

Artificial Intelligence is the generally accepted premier international forum for the publication of results of current research in the artificial intelligence field [34]. It includes foundational and applied papers describing mature work involving computational accounts of aspects of intelligence [34].

2) **Big Data** [34]

Big Data provides the research in collecting, analyzing, and disseminating vast amounts of data, including data science, big data infrastructure and analytics, and pervasive computing [34].

3) **Computational Statistics & Data Analysis by International Association for Statistical Computing (IASC)** [34]

Computational Statistics & Data Analysis (CSDA), the official journal of the International Association of Statistical Computing (IASC), is an international journal that aims to the dissemination of methodological research and applications in the areas of computational statistics and data analysis [34].

4) **International Journal of Business Intelligence and Data Mining** [34]

IJBIDM provides a forum for the developments, research, and innovation of business intelligence, data analysis and mining [34].

5) **Big Data Research** [34]

The journal provides a high-quality big data forum for researchers, practitioners, and policymakers from the many different communities [34].

Professional/Trade Journals

1) **Analytics Insight Magazine** [35]

Analytics Insight is a technology platform that focuses on insights, trends, and opinions from the world of data-driven technologies [35].

2) **Emerj** [35]

Emerj Artificial Intelligence Research is a platform that provides the industry source for market research to give decision-makers the insight, AI technology procurement and strategic planning around AI [35].

3) **Dataversity** [35]

Dataversity produces educational resources for business and Information Technology (IT) professionals on the uses and management of data [35].

4) **Datafloq** [35]

People can go through high-quality articles, and find big data and technology vendors via Datafloq [35]. Datafloq provides information, insights, and opportunities to drive innovation with big data, blockchain, artificial intelligence and other technologies [35].

5) **Dataconomy** [35]

Dataconomy is the leading portal for news, events and expert opinion from the world of data-driven technology [35].

To Do Research

To do research in the data analytics domain, the processes usually involve asking an interesting question, getting the data, exploring the data, modeling, evaluating and analyzing, communicating and visualizing the results. ***This brochure will further discuss methods of the collection of data, modeling, and evaluation.***

Modeling: Research Methods to Answer Questions

As mentioned in the introduction, data analysts mainly address problems via the knowledge of statistics, operations research, data mining, machine learning, deep learning, database, big data, visualization, etc. The specific functions and introductions are showing as follows:

- **Statistics knowledge** not only is the foundation of many other domains but also provides data analysts with correlation analysis models, statistical prediction models, statistical test models, etc. [15] [16]. The specific Moreover, the statistical test is one of the most important approaches to evaluate data analytics models [16].
- **Operations research** is a mathematical branch that mainly investigates the optimization problems [17], such as what the shortest path of packages delivery is [36], what the most appropriate number of security check channels in airports is [37], and what the solution of maximum return is [23].
- **Data mining** mainly provides several algorithms to extract valuable information from given datasets and make prediction [2], such as anomaly detection [21], association analyzation [24], clustering [20], regression [19], classification [2], etc. For example, K-Means is a clustering algorithm [20]; linear regression and logistic regression are regression algorithms [19]; decision tree, SVM, and K-NN are classification algorithms [2].

- **Machine learning** combines the technology of data mining and other algorithms to provide more advanced data mining technologies [2]. In addition, deep learning, the development of machine learning, is based on the neural network [2]. The advantage of it is excellent performance on large scale data [2].
- **The database technique** serves the storage and access of data to address the analyzation of complex data [4]. In addition, due to more complicated demands of data analyzation, data warehouse and data lake were come up with [4].
- With the dramatic growth of data nowadays, **big data technique** provides us with more opportunities and challenges on data science domain [5]. What needs to be pointed out is that big data is mainly the automatic records on the Internet or sensors rather than the pre-collection of data with a specific purpose [5]. In addition, the specific techniques of big data are shown in the section of pressing issues.
- **Visualization technique** functions as exploratory data analysis, error detection, and communication [9]. To be specific, exploring data by visualization is the first step to analyze the problem; visualization can reveal the hidden information of statistics analyzation to detect errors; the visualization of results is an efficient way to convey information [9].

To Do Research

To do research in the data analytics domain, the processes usually involve asking an interesting question, getting the data, exploring the data, modeling, evaluating and analyzing, communicating and visualizing the results. ***This brochure will further discuss methods of the collection of data, modeling, and evaluation.***

Evaluation: What is Considered Credible Evidence of the Research?

After building a strong model, how can we convince others to accept the results we have analyzed? Hence, validating our models via scientific methods is as important as modeling. Though it is surprisingly hard to have a precise evaluation, it can provide an essential interpretation of the model and results [1].

Here show basic thoughts of evaluation in the data science domain.

- **As for large enough data**, we can separate the dataset to training subset for training the model, test subset for testing the model, and validation subset for confirming the performance of the final model before it goes into production [1]. **As for small samples**, we can use Cross-validation to give an evaluation report [1]. In addition, some experiments are not required to make the dataset separate.

- **The next step is to evaluate the specific model**. First of all, the statistical test is the foundation of many other evaluation methods. The statistical summary and statistical test can show what the distribution of results is, what the statistical description of results is, whether observations on data is significance, and how much we could trust the result [15] [16]. Specific methods are mean, standard deviation, T-Test, Friedman-Test, Chi-Squared Test, Pearson's Correlation Coefficient, etc. [16]. Based on statistics, different problems have more specific validation algorithms. For instance, as for classification problems, we can calculate the Accuracy, Precision, Recall, and F-Score of results based on confusion matrixes [1]. In addition, Receiver-Operator Characteristic (ROC) Curves and statistical summary are also used in evaluating categories prediction [1].

As you can see, different analyzation models have different approaches to fit and evaluate them. We need to distinguish them correctly and choose the appropriate ways without bias [1]. In addition, it is very easy for data scientists to fool themselves and others [1]. For example, a validation method that has the highest score can be selected with a personal bias to praise the model [1]. They could have never thought about reasons why other validation methods have poor performance, even have never used comprehensive methods to give an evaluation.

Hence, credible evidence not only simply involves the above methods but also requires our efficient design of the validation model.

Writing

Genres in the Data Science Domain

Popular articles, trade articles, short reports, academic articles, technical reports, and case studies are the main six writing genres in the data analysis domain. They function in different sceneries, use different tones, and have different audiences and structures. Even articles structures in one genre may have differences according to the specific contexts. Specifically:

1) **Popular articles** use the easy understanding language and an informal tone to guide the public to know about some scientific research.
2) **Trade articles** informally discuss some ideas, investigations, and analyzations in one industry. They may involve some professional words and knowledge which are difficult to understand for the general public.
3) **Short reports** are usually written by researchers to give a concise description and introduction of a proposal, an experiment, a progress report, or a conference. Their audiences usually are the people who ask or feel interested in the report.
4) **Technical reports** mainly focus on the comprehensive description of the research in detail. People who work in this field read them.
5) **Academic articles**, a more formal and professional genre than others, are the final reports of big projects and researches that will be published.

Particularly, academic articles are the most helpful for data analysts because they not only have the most formal writing style but also propose novel solutions in the data science domain. Students can enhance their both academic skills and writing abilities.

Here further introduces the structure of academic articles. This genre mainly consists of five elements which are abstract, introduction, methods, results, discussion and reference.

- **Abstract** contains the brief introduction of the research object, exigence, purpose, methods and main findings.
- **Introduction** part will further introduce the general trend, exigence, purpose, and relevant researches.
- **Methods** part, the main body of an academic article, provides the specific detail introduction of research approaches.
- **Results** part shows the accomplishment of the proposed methods and relevant evaluation of the results.
- **Discussion** part will not only restate the research object, exigence, and purpose, but also summarize the methods and main findings. In addition, it will discuss the relevance of the research, including the significance and future works.
- **Reference** part will list all referenced articles in the research.

Writing

Differences between a Scholarly Article and a Technical/Professional Text

As shown in the introduction of academic articles, a scholar article has a more formal writing style on structure. Though sometimes the specific research will modify the given template according to the context, those six elements cannot be dropped. However, the technical text focus on the clear delivery of what the research has done and what the specific experiment details are. In addition, the technical text does not require a list of references like the academic article. Moreover, the language style and tone of scholar articles are more formal and accurate than those of technical text.

Reference Sources

Writing

- **Writing Center at GMU** provides the service of tutoring, workshops, and ESL writing supports. https://writingcenter.gmu.edu/

- This website offers **the introduction of different document types in the engineering domain**. https://ecp.engineering.utoronto.ca/resources/online-handbook/types-of-documents/

- **The InfoGuides of Data Analytics Engineering discipline**: https://infoguides.gmu.edu/DAEN

Tutorial

- **The InfoGuides of data collection**: https://dsc.gmu.edu/

- Use Mason ID to take the free online **LinkedIn Learning courses**, such as R, Python, etc. https://lil.gmu.edu/

Community

- **Kaggle**, the largest and most diverse data community in the world, holds a great deal of online machine learning competitions and provides enormous open datasets used in data science. https://www.kaggle.com/
- **Reddit** is an American social news aggregation, web content rating, and discussion website. https://www.reddit.com/r/datascience/
- **Data Science Central** is a community that provides big data practitioners with the industry's online resource, including data analytics, data integration, data visualization, etc. https://www.datasciencecentral.com/

References

[1] S. S. Skiena, "Mathematical Models," in *The Data Science Design Manual*, vol. 7, New York, Springer Berlin Heidelberg, 2017.

[2] S. S. Skiena, "Machine Learning," in *The Data Science Design Manual*, vol. 11, New York, Springer Berlin Heidelberg, 2017.

[3] NIST Big Data Public Working Group Definitions and Taxonomies Subgroup, "NIST Big Data Interoperability Framework: Volume 1, Definitions," National Institute of Standards and Technology, 2015.

[4] Panoply, "Modern Data Management: Next Generation Data Tools Eliminate Data Maintenance," [Online]. Available: https://panoply.io/data-warehouse-guide/.

[5] S. S. Skiena, "Big Data: Achieving Scale," in *The Data Science Design Manual*, vol. 12, New York, Springer Berlin Heidelberg, 2017.

[6] "Internet Live Stats," [Online]. Available: https://www.internetlivestats.com/.

[7] "Twitter Usage Statistics," [Online]. Available: https://www.internetlivestats.com/twitter-statistics/.

[8] "Google Search Statistics," [Online]. Available: https://www.internetlivestats.com/google-search-statistics/.

[9] S. S. Skiena, "Visualizing Data," in *The Data Science Design Manual*, vol. 6, New York, Springer Berlin Heidelberg, 2017.

[10] C. Peter, Elements of Financial Risk Management, 2011.

[11] T. C and W. J, "Using neural network ensembles for bankruptcy prediction and credit scoring," *Expert Systems with Applications,* vol. 34, no. 4, pp. 2639-2649, 2008.

[12] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu and I. Stancu, "Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 543, no. 2-3, pp. 577-584, 2005.

[13] National Cancer Institute, "Cancer Statistics," National Cancer Institute, 2018. [Online]. Available: https://www.cancer.gov/about-cancer/understanding/statistics.

[14] T. B. Murdoch and A. S. Detsky, "The Inevitable Application of Big Data to Health Care," *JAMA,* vol. 309, no. 13, p. 1351, 2013.

[15] S. S. Skiena, "Statistical Analysis," in *The Data Science Design Manual*, vol. 5, New York, Springer Berlin Heidelberg, 2017.

[16] E. Marshall, The Statistics Tutor’s Quick Guide to Commonly Used Statistical Tests.

[17] A. Sharma, Operations Research, Mumbai, INDIA: Global Media, 2018.

References

[18] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in *Sigmod Record*, 1993.

[19] S. S. Skiena, "Linear and Logistic Regression," in *The Data Science Design Manual*, vol. 9, New York, Springer Berlin Heidelberg, 2017.

[20] S. S. Skiena, "Distance and Network Methods," in *The Data Science Design Manual*, vol. 10, New York, Springer Berlin Heidelberg, 2017.

[21] S. S. Skiena, "Data Munging," in *The Data Science Design Manual*, vol. 3, New York, Springer Berlin Heidelberg, 2017.

[22] R. . W. Hornung, A. L. Greife, L. T. Stayner and N. K. Steenland, "Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study," *American Journal of Industrial Medicine,* vol. 25, no. 6, pp. 825-836, 1994.

[23] M. Rouse, "operations research (OR)," 2019. [Online]. Available: https://whatis.techtarget.com/definition/operations-research-OR.

[24] A. Rai, "An Overview of Association Rule Mining and its Applications," 2019. [Online]. Available: https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/.

[25] S. Bandyopadhyay and U. Maulik, "Genetic clustering for automatic evolution of clusters and application to image classi#cation," *Pattern Recognition,* vol. 35, no. 6, pp. 1197-1208, 2002.

[26] A. Zimek and E. Schubert, "Outlier Detection," in *Encyclopedia of Database Systems*, New York, Springer, 2017, pp. 1-5.

[27] Y. Kou, C.-T. Lu, S. Sinvongwattana and Y.-P. Huang, "Survey of fraud detection techniques," in *IEEE International Conference on Networking, Sensing and Control*, 2004.

[28] Digital Scholarship Center at GMU, "Digital Scholarship Center," [Online]. Available: https://dsc.gmu.edu/.

[29] Association for Computing Machinery, "The ACM Digital Library," [Online]. Available: https://dl.acm.org/dl.cfm?coll=portal&dl=ACM.

[30] EBSCO Industries, "Computers & Applied Sciences Complete," [Online]. Available: http://web.a.ebscohost.com/ehost/search/advanced?vid=0&sid=923b61b5-a8d7-4fce-a1d6-ec47156b0acd%40sdc-v-sessmgr02.

[31] IEEE, "IEEE Xplore Digital Library," [Online]. Available: https://ieeexplore.ieee.org/search/advanced.

[32] Elsevier B.V., "ScienceDirect," [Online]. Available: https://www.sciencedirect.com/.

[33] Springer Nature Switzerland AG., "Springer Link," [Online]. Available: https://link.springer.com/.

[34] B. Gupta, "10 ESSENTIAL ACADEMIC JOURNALS FOR DATA SCIENTISTS," [Online]. Available: https://analyticsindiamag.com/10-essential-academic-journals-data-scientists/.

[35] IndustryWired, "Top 10 Big Data and Artificial Intelligence Magazines and Publications," 26 June 2019. [Online]. Available: https://industrywired.com/top-10-big-data-and-artificial-intelligence-magazines-and-publications/.

References

[36] D. Kim, C. Barnhart and K. Ware, "Multimodal Express Package Delivery: A Service Network Design Application," *Transportation Science,* vol. 4, no. 391-407, p. 1999, 33.

[37] A. G. d. Barros and D. D. Tomber, "Quantitative analysis of passenger and baggage security screening at airports," *Journal of Advanced Transportation,* vol. 44, no. 1, pp. 171-193, 2007.

[38] S. S. Skiena, "Mathematical Preliminaries," in *The Data Science Design Manual*, vol. 2, New York, Springer Berlin Heidelberg, 2017.

[39] S. S. Skiena, "Big Data: Achieving Scale," in *The Data Science Design Manual*, vol. 12, New York, Springer Berlin Heidelberg, 2017.

[40] Z. Xi, H. Wang, Y. Hao, J.-M. Lien and I.-S. Choi, "Compact Folding of Thick Origami via Stacking," Department of Computer Science, George Mason University, 2017.

[41] University of Toronto, Faculty of Applied Science and Engineering, "Engineering Communicaition Program," University of Toronto, [Online]. Available: https://ecp.engineering.utoronto.ca/resources/online-handbook/types-of-documents/lab-reports/.