InfoMINDS: An Interdisciplinary Framework for Leveraging Data Science upon Big Data in Surface Mining Industry

Vitor Afonso Pinto[1][a] and Fernando Silva Parreiras[2][b]

[1]*Technology Department, Operational Technology for Base Metals South Atlantic, Vale, Carajas, Para, Brazil*

[2]*Laboratory for Advanced Information Systems, FUMEC University, Rua do Cobre, Belo Horizonte, Brazil*

Keywords: Data Science, Big Data, Framework, Mining Industry.

Abstract: Intending to be more and more data-driven, companies are leveraging data science upon big data initiatives. However, to reach a better cost-benefit, it is important for companies to understand all aspects involved in such initiatives. The main goal of this paper is to provide a framework that allows professionals from the mining industry to accurately describe data science upon big data. The following research question was addressed: *"Which essential components characterize an interdisciplinary framework for data science upon big data in mining industry?"*. To answer this question, we will extend OntoDIVE ontology to create a framework capable of explaining aspects involved in such initiatives for the mining industry. As a result, this paper will present InfoMINDS - A Framework for Data Science upon Big Data Relating People, Processes and Technologies on Mining Industry. This paper will contribute to leveraging data science initiatives upon big data allowing application of OntoDIVE on real-case scenarios in mining industry.

1 INTRODUCTION

Data Science can be defined as an approach to extract worthy insights from low-value data. Big Data can be defined as an integrated ecosystem of technologies performing formal roles with the purpose to create technical conditions for the delivery of value-added applications based on data. Data science upon big data is seen by organizations as a tool to improve operational efficiency though it has strategic potential, drive new revenue streams and gain competitive advantages (Sivarajah et al., 2017).

There is a considerable literature addressing major concepts related to data science and big data. Some studies proposed ways to characterize data science upon big data using concepts of volume, velocity, variety, validity, veracity, variability, visibility, verdict and value (Sharma, 2017; Corea, 2016; Addo-Tenkorang and Helo, 2016). Other studies proposed ways to group big data technologies (Bari et al., 2014; Ahlemeyer-Stubbe and Coleman, 2014; Murthy et al., 2014). Other studies proposed ways to explain data science processes (Corea, 2016; Maimon and Rokach, 2010; Abbott, 2014; Takurta et al., 2017).

Additionally, a range of literature exists suggesting an interdisciplinary approach as a success factor for data science initiatives (Corea, 2016; Forte, 2015; Cady, 2017; Luis, 2017). Other studies presented results generated by data science initiatives that involved multiple knowledge areas (Xu et al., 2014; Fisher et al., 2017; Roy et al., 2017; Capalbo et al., 2017; Zhou et al., 2016; Arias and Bae, 2016; Santoro et al., 2018; Hurwitz et al., 2015; Van Der Aalst, 2016; Zheng et al., 2015; Talón-Ballestero et al., 2018; Balliu et al., 2016; Lei et al., 2016; Maciejewski, 2017; Lu and Li, 2017; Stoet and Geary, 2018; Seele, 2017; Lubchenco and Grorud-Colvert, 2015).

Even with all the progress that has been made, mining industry is still grappling with how to capture insights that are not obvious. As an example, mining personnel not always understand how data science initiatives are conducted in other industries. Another aspect is that mining companies not always have expertise to define and deploy big data technological ecosystems. Although big data comprises technologies performing formal roles, technologies may vary among organizations and each technology should be minutely chosen to avoid loss of effectiveness. Thus, there is a risk of receiving biased advisory from consultants who try to push technologies based on their own interests or limitations.

[a] https://orcid.org/0000-0002-2731-0952
[b] https://orcid.org/0000-0002-9832-1501

Pinto, V. and Parreiras, F.
InfoMINDS: An Interdisciplinary Framework for Leveraging Data Science upon Big Data in Surface Mining Industry.
DOI: 10.5220/0010484107840791
In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 2*, pages 784-791
ISBN: 978-989-758-509-8

Considering the lack of concepts as a major contributing factor for preventing the leverage of data science initiatives upon big data, it is crucial to explain all concepts related to this kind of initiative and the interactions between and among them. The main goal of this paper is to propose a conceptual framework that allows professionals from the mining industry to accurately describe data science upon big data.

Our approach differs from others as we intend to create common ground so that data science initiatives upon big data can be fully understood by professionals from any area of knowledge. In order to contribute to the body of knowledge, this paper is more interested in the general idea or conception behind data science initiatives upon big data rather than any individual instance of those initiatives. In this context, the following research question was addressed: *"Which essential components characterize an interdisciplinary framework for data science upon big data in mining industry?"*.

By addressing this research question, in this paper we propose InfoMINDS which is a conceptual framework that organizes practices commonly applied during data science initiatives upon big data, considering a comprehensive and end-to-end perspective. This framework may contribute either to clarification of concepts or the explanation of interactions between and among them. InfoMINDS can also be considered as a foundation upon which mining companies can build policies, standards, rules, procedures, methodologies or any other artifact in order to leverage data science initiatives upon big data. This paper is structured as follows: Section 1 presents the context of this work. Section 2 presents methods of research. Section 3 presents results that are discussed in Section 4. Section 5 concludes the paper.

2 METHODS

A conceptual framework may be defined as an end result of bringing together a number of related concepts to explain or predict a given event or to give a broader understanding of the phenomenon of interest (Imenda, 2014). It is a visual presentation of key variables, factors or concepts and their relationship among each other which have been or have to be studied (Miles and Huberman, 1994). The main purpose of a conceptual framework is to bring focus on the content and to act as a link between literature, methodology and results. We decided to build a conceptual framework as it provides understanding, rather than offering a theoretical explanation (Jabareen, 2009).

In order to guarantee completeness and correctness of InfoMINDS Framework, we decided to implement it using OntoDIVE Ontology, presented in (Pinto and Parreiras, 2020). In practical terms, we created individuals on class **Frameworks** to represent either the InfoMINDS framework itself and each one of its dimensions. We also created individuals on class **Processes** to represent each of InfoMINDS processes. As this study is focused on mining industry, we created individuals on Class **Frameworks** to represent either the mining industry and each one of its production phases (mining and mineral processing).

3 RESULTS

3.1 InfoMINDS Structure

InfoMINDS is structured in twenty-five processes that are grouped in five dimensions. Processes derived from existing multidisciplinary literature. Dimensions derived from ICT business processes: Plan, Build, Run, Enable and Manage. InfoMINDS is designed to be flexible as each initiative is unique. Figure 1 presents InfoMINDS Framework. Next subsections present details on dimensions and processes. We made InfoMINDS available at GitHub [1] to encourage its usage in future studies.

3.1.1 Dimension A: Plan

Dimension A (Plan). includes processes that aim at understanding and defining the goals of end users and the environment in which data science initiative will take place. Table 1 presents the five processes included in this dimension along with their major goals.

3.1.2 Dimension B: Build

Dimension B (Build). includes processes to selecting, preprocessing, transforming data and also modeling and evaluating data applications. Table 2 presents the five processes included in this dimension along with their major goals.

3.1.3 Dimension C: Run

Dimension C (Run). includes processes to deploy, manage and monitor data application and the outcomes generated by them. Table 3 presents the three processes included in this dimension along with their major goals.

[1] https://github.com/tecladista1/InfoMINDS

Table 1: Processes from Dimension A (Plan).

Process	Major Goals
Acquire Interdisciplinary Team	To provide an interdisciplinary team capable of executing all required activities to achieve results proposed by data science initiatives. This process focus on explanation of some functions that need to be performed so that a data science initiative can be successful.
Understand Business Context	This process involves understanding the goals of the end-user in terms of what is expected from data science initiative, considering all existing constraints and requirements. *Business Process Modeling* is an approach that could be used to facilitate this process.
Define Appropriate Paradigms	To choose appropriate paradigms for conducting a data science initiative considering the current scenario. Depending on the expected results a different paradigm may be chosen: agile, waterfall, among others.
Define Technological Ecosystem	To define a technological ecosystem to support a data science initiative. Although Big Data comprises technologies performing formal roles, the technologies chosen to perform each role may vary among organizations or even among initiatives.
Determine Initiative Readiness	To determine the maturity of people, processes and technologies to conduct a data science initiative upon big data. This process intends to clarify if all conditions are set to starting a data science initiative upon big data.

*Source: Authors.

3.1.4 Dimension D: Enable

Dimension D (Enable). includes processes to address computing infrastructure, procurements, cybersecurity and other enabling processes. Table 4 presents processes included in this dimension and their goals.

Table 2: Processes from Dimension B (Build).

Process	Major Goals
Perform Data Selection	This process consists in identifying data sources, acquiring, integrating and transferring data. Data need to be consistently aggregated from different sources of information, and integrated with other systems and platforms.
Perform Data Preprocessing	Much of the raw data contained in databases is unpreprocessed, incomplete, and noisy. The main goal of this process is to treat outliers, inconsistent values, missing values, redundant fields and obsolete fields.
Perform Data Transformation	The main goal of this process is to transform or consolidate data so that the resulting data science processes may be more efficient. Data transformation comprehends transformation, dimension reduction and discretization of data.
Perform Data Modeling	The main goal of this process is to create a model based on initial hypothesis, exploratory data analysis, classification, clustering, among others. Models can be equations linking quantities that we can observe or measure. They can also be a set of rules.
Perform Model Evaluation	The main goal of this process is to perform validation and verification tests of data application. Model evaluation is the process of assessing a property or properties of a model in terms of its structure and data inputs so as to determine whether or not the results can be used in decision-making

*Source: Authors.

3.1.5 Dimension E: Manage

Dimension E (Manage)., includes processes to address strategical processes, portfolio management, risks, among others. Table 5 presents processes included in this dimension and their goals.

Table 3: Processes from Dimension C (Run).

Process	Major Goals
Deploy Data Application	The main goal of this process is to deploy authorized version of data application in a production environment.
Manage Data Application	The main goal of this process is to manage data application, including incidents, problems, changes, among others.
Monitor Data Application Outcomes	The main goal of this process is to capture outcomes provided by data application.

*Source: Authors.

Table 4: Processes from Dimension D (Enable).

Process	Major Goals
Develop and Manage Team	To manage and develop team allocated to a particular data science initiative.
Deploy and Improve Computing Infrastructure	To deploy technologies for data creation, acquisition, transmission, ingestion, storage, pre-processing, data modeling, among others.
Manage and Control Procurements	To provide contracts with external vendors and partners.
Implement and Monitor Data Governance	To implement and monitor the maturity level of data governance practices.
Implement and Monitor Information Security	To implement and monitor policies and routines to prevent or mitigate risks related to information security.
Manage and Control Budget	To manage and control economic and financial budgets.

*Source: Authors.

3.2 InfoMINDS Application

In this section we present how we applied InfoMINDS Framework to data science initiatives upon big data on the mining industry. We used InfoMINDS to organize activities of four data science initiatives: two initiatives focused on **operations** and two focused on **maintenance**. Initiatives followed the sequence indicated by InfoMINDS Dimensions: Plan, Build, Run, Enable and Manage. To ensure correctness and completeness of this stage, we used OntoDIVE ontology, presented in (Pinto and Parreiras, 2020) to represent all elements involved on initiatives.

Table 5: Processes from Dimension E (Manage).

Process	Major Goals
Provide Strategy Alignment	To share business strategy with all data science initiatives upon big data.
Manage Overall Portfolio	To manage overall portfolio of either projects or services..
Manage and Control Risks	To identify, assess, manage and control risks.
Manage and Control Resources	To manage and control non-financial resources available for all initiatives.
Manage Benefits Realization	To manage benefits realization for all the initiatives.
Share Knowledge and Information	To share knowledge and information about previous initiatives.

*Source: Authors.

3.2.1 Implementation of Dimension A: Plan

Following guidelines of Processes **A.1 Acquire Interdisciplinary Team**, **A.2 Understand Business Context**, **A.3 Define Appropriate Paradigm**, **A.4 Define Technological Ecosystem** and **A.5 Determine Initiative Readiness**, interdisciplinary teams were allocated to each initiative. Next, each squad was assigned to a real business problem and teams chose the most appropriate paradigm for the initiative under their responsibility. Next, teams searched OntoDIVE Ontology to find existing technologies previously implemented by mining processes specialists. At the end, teams decided to move all initiatives to the next stage after considering their readiness.

3.2.2 Implementation of Dimension B: Build

Following guidelines of Processes **B.1 Perform Data Selection**, **B.2 Perform Data Preprocessing**, **B.3 Perform Data Transformation**, **B.4 Perform Data Modeling** and **B.5 Perform Model Evaluation**, all teams performed activities to select, preprocess and transform data from available data sources. Next, data models were created and evaluated. Actions performed in this stage were different as each initiative had a different starting point. While some initiatives were focused on rolling out existing applications to a different location, other initiatives had to build predictive models and applications from the scratch.

3.2.3 Implementation of Dimension C: Run

Following guidelines of Processes **C.1 Deploy Data Application**, **C.2 Manage Data Application**, **C.3 Monitor Data Application Outcomes**, all teams performed activities to deploy and manage data applications besides monitoring outcomes of these data applications. We created individuals on Class **Outcomes** to represent data applications deployed in this stage. These data applications were linked to roles in the big data technological ecosystem through Object Property **haveRole**.

3.2.4 Implementation of Dimension D: Enable

Following guidelines of Processes **D.1 Develop and Manage Team**, **D.2 Deploy and Improve Computing Infrastructure**, **D.3 Manage and Control Procurements**, **D.4 Implement and Monitor Data Governance**, **D.5 Implement and Monitor Information Security** and **D.6 Manage and Control Budget**, team "TE00 – Squad Shared Services" oversaw people development, management of computing infrastructure, management of budget and procurements and monitoring of data governance and information security. Initiatives started exchanging files and as the time went on, improvements were made to enable real-time acquisition of data.

3.2.5 Implementation of Dimension E: Manage

Following guidelines of Processes **E.1 Provide Strategy Alignment**, **E.2 Manage Overall Portfolio**, **E.3 Manage and Control Risks**, **E.4 Manage and Control Resources**, **E.5 Manage Benefits Realization** and **E.6 Share Knowledge and Information**, this dimension comprehends processes to either start or finish initiatives. To start the initiatives analyzed by this study, first they were authorized. Then, all financial and non-financial resources were made available. We used object property **supportedBy** to link initiatives to people, processes and technologies.

3.3 InfoMINDS Examples

This section presents outcomes generated by initiatives analyzed in this paper. Examples of this section were included to illustrate the potential of InfoMINDS Framework.

3.3.1 Data Science for Mining Operations

This initiative was created to improve energy efficiency in mining operations, by reducing fuel consumption. In practical terms, this initiative intended to address the following business problem: *In what extension the fuel consumption on coal mining operations is affected by other variables?*. Table 6 presents outcomes of this initiative in order to illustrate InfoMINDS capabilities.

3.3.2 Data Science for Mineral Processing

Mineral processing includes size reduction and enrichment of minerals. This initiative was created to improve coal marketability, by ensuring values of yield above 30% and ash below 11.2%. In practical terms, this initiative intended to address the following business problem: *In what extension ash and yield are affected by other variables of coal mineral processing?*. Table 7 presents outcomes of this initiative in order to illustrate InfoMINDS capabilities.

3.3.3 Data Science for Mining Maintenance

Industrial maintenance intends to guarantee availability and reliability for facilities and equipment. This initiative was created to improve asset management, by extending lifetime of mining trucks. In practical terms, this initiative intended to address the following question: *In what extension the lifetime of mining trucks are affected by other process variables?*. Table 8 presents outcomes of this initiative in order to illustrate InfoMINDS capabilities.

3.3.4 Data Science for Plant Maintenance

Plant maintenance seeks for optimum availability, optimum operating conditions, maximum utilization of maintenance resources, optimum equipment life, minimum spares inventory and ability to react quickly. This initiative was created to improve asset management, by reducing tearing on conveyor belts. In practical terms, this initiative intended to address the following question: *In what extension tearing of conveyor belts are related to other process variables?*. Table 9 presents outcomes of this initiative in order to illustrate InfoMINDS capabilities.

4 DISCUSSION

InfoMINDS is a conceptual framework that organizes practices commonly applied to design, build and maintain data applications considering a comprehensive and end-to-end perspective. InfoMINDS creates a common vocabulary allowing processes for development and maintenance to be reused and shared amongst industries from different segments.

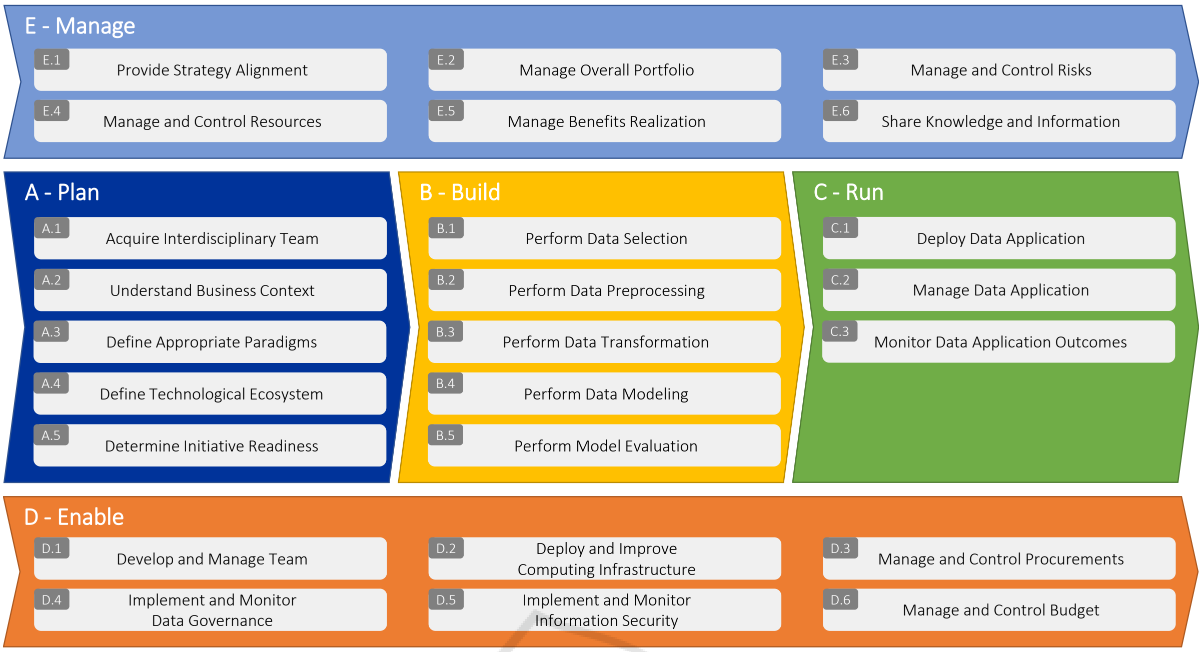

Figure 1: InfoMINDS: Framework Overview.

*Source: Authors.

Table 6: Examples of Data Science for Mining Operations.

Predictive Application	A predictive application was built to identify processes variables that could influence fuel consumption based on routes, conditions of roads and inclination angles of equipments.
Prescriptive Application	A prescriptive application was built and deployed into the fleet management system which is used by operators to receive information from dispatch controllers.
Descriptive Application	A descriptive application was built and deployed to facilitate finding deviations to recommendations of predictive application.

*Source: Authors.

InfoMINDS may contribute either to clarification of concepts or to the explanation of interactions between and among them. It can also be considered as a foundation upon which mining companies can build policies, standards, rules, procedures, methodologies or any other artifact in order to leverage data science initiatives upon big data.

InfoMINDS allows professionals from any knowledge area to understand major processes and activities related to data science initiatives upon big data as well as the processes and activities related to the management, maintenance and support of data applications. All initiatives analyzed in this study were led by personnel with no previous knowledge about big data technologies or data science management.

InfoMINDS contributes either to clarification of concepts or to the explanation of interactions between and among them. Mining industry can benefit from InfoMINDS as transaction technologies, previously implemented by mining processes specialists, generate large volumes of scattered data and integrated analyses of those data may be used as a tool for improving operational efficiency of the industry. InfoMINDS takes all existing technologies into consideration as they have the potential to be used as data source for data science initiatives. Besides that, InfoMINDS helps mining industry to share best practices between different sites. Data applications showed in this paper could be rolled out to similar operations, leveraging data science upon big data.

5 CONCLUSION

InfoMINDS framework is a conceptual framework designed to create common ground so that data science initiatives upon big data can be fully understood by professionals from any area of knowledge. It has twenty-five processes grouped into five dimensions and contributes to leveraging data science initiatives

Table 7: Examples of Data Science for Mineral Processing.

Predictive Application	A predictive application was built and deployed to identify processes variables that could influence results of *ash* and *yield*. This application was designed to identify the physical type of coal being processed and automatically suggest parameters so that results of *ash* and *yield* could be inside an expected range.
Prescriptive Application	A prescriptive application was built and deployed into the system used by operators to determine setups for processes variables. This prescriptive application brought tangible results as operators were able to see the distance between their setups and the optimal ranges.
Descriptive Application	A descriptive application was built and deployed to facilitate finding deviations to recommendations of predictive application. This application was used to monitor adherence by operator, by shift, and so on.

*Source: Authors.

Table 8: Examples of Data Science for Mining Maintenance.

Predictive Application	A predictive application was built and deployed to identify variables that could influence lifetime of mining trucks. This application analyzes fueling data, haul truck telemetry, engineering parameters and laboratory results to recommend the sequence of scheduled maintenance.
Prescriptive Application	A prescriptive application was built and deployed into the maintenance workshop. This application is focused on presenting trucks with lifetime lower than expected and suggests the components to be replaced to extend the lifetime of equipment.
Descriptive Application	A descriptive application was built and deployed in order to facilitate finding deviations to recommendations of predictive application. This application is used to monitor components suggested to be replaced.

*Source: Authors.

Table 9: Examples of Data Science for Plant Maintenance.

Predictive Application	A predictive application was built and deployed to identify variables that could quickly detect tearing of conveyor belts. After analyzing different process variables, a single process variable was identified as capable of indicating the start of a tearing event.
Prescriptive Application	A prescriptive application was built and deployed into the plant controller to automatically stop conveyor belts whenever requested by predictive application.

*Source: Authors.

upon big data in mining industry. It can be considered as a foundation upon which mining companies can build policies, standards, rules, procedures, methodologies or any other artifact, in order to leverage data science. It also may help mining industry professionals to draw parallels between data science results for a different domain to their own domain.

InfoMINDS confirmed its capability of explaining interactions between people, processes and technologies in the context of data science upon big data on mining industry. The framework confirmed its contribution to the clarification of concepts and terminologies related to either data science or big data in mining industry. In this study, all initiatives were led by personnel with no previous knowledge about data science. Still, consistent data science results were achieved. InfoMINDS showed it can enlarge possibilities of data science applications.

This study has several limitations. Firstly, InfoMINDS is based on OntoDIVE ontology and inherits its limitations. Besides that, InfoMINDS was conceived as a conceptual framework. Assuming that different researchers may approach a single phenomenon using different perspectives, it is possible that they might end up with different conceptual frameworks as final result. Another limitation is the fact InfoMINDS was applied in data science initiatives of a single mining company. Future works could apply InfoMINDS on more real-case scenarios to collect insights and thoughts of more people. Future works could also implement a system based on OWL file generated by Protegé.

REFERENCES

Abbott, D. (2014). *Applied predictive analytics: principles and techniques for the professional data analyst*. Wiley, Indianapolis (IN).

Addo-Tenkorang, R. and Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering*, 101:528 – 543.

Ahlemeyer-Stubbe, A. and Coleman, S. (2014). *A practical guide to data mining for business and industry*. John Wiley & Sons.

Arias, M. B. and Bae, S. (2016). Electric vehicle charging demand forecasting model based on big data technologies. *Applied energy*, 183:327–339.

Balliu, A., Olivetti, D., Babaoglu, Ö., Marzolla, M., and Sîrbu, A. (2016). A big data analyzer for large trace logs. *Computing*, 98(12):1225–1249.

Bari, A., Chaouchi, M., and Jung, T. (2014). *Predictive Analytics for Dummies*. Wiley, [s.l].

Cady, F. (2017). *The data science handbook*. John Wiley & Sons, Inc., Hoboken, NJ.

Capalbo, S. M., Antle, J. M., and Seavert, C. (2017). Next generation data systems and knowledge products to support agricultural producers and science-based policy decision making. *Agricultural systems*, 155:191–199.

Corea, F. (2016). *Big data analytics: a management perspective*, volume 21. Springer, [s.l].

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., et al. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, 53(4):2618–2626.

Forte, R. M. (2015). *Mastering predictive analytics with R: master the craft of predictive modeling by developing strategy, intuition, and a solid foundation in essential concepts*. Packt Publishing, Birmingham, UK.

Hurwitz, J., Kaufman, M., and Bowles, A. (2015). *Cognitive Computing and Big Data Analytics*. Wiley Publishing, [s.l], 1st edition.

Imenda, S. (2014). Is there a conceptual difference between theoretical and conceptual frameworks? *Journal of Social Sciences*, 38(2):185–195.

Jabareen, Y. (2009). Building a conceptual framework: philosophy, definitions, and procedure. *International journal of qualitative methods*, 8(4):49–62.

Lei, Y., Jia, F., Lin, J., Xing, S., and Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5):3137–3147.

Lu, H. and Li, Y. (2017). *Artificial intelligence and computer vision*. Springer.

Lubchenco, J. and Grorud-Colvert, K. (2015). Making waves: The science and politics of ocean protection. *Science*, 350(6259):382–383.

Luis, T. (2017). *Data mining with R: learning with case studies*. CRC Press, Taylor and Francis Group, Boca Raton.

Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, 83(1_suppl):120–135.

Maimon, O. and Rokach, L. (2010). *Data mining and knowledge discovery handbook*. Springer, New York.

Miles, M. B. and Huberman, A. M. A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. Sage Publications, Thousand Oaks, Calif., 2nd ed. edition.

Murthy, P., Bharadwaj, A., Subrahmanyam, P. A., Roy, A., and Rajan, S. (2014). Big data taxonomy.

Pinto, V. A. and Parreiras, F. S. (2020). Ontodive: An ontology for representing data science initiatives upon big data technologies. In *ICEIS (1)*, pages 42–51.

Roy, S., Ray, R., Roy, A., Sinha, S., Mukherjee, G., Pyne, S., Mitra, S., Basu, S., and Hazra, S. (2017). Iot, big data science & analytics, cloud computing and mobile app based hybrid system for smart agriculture. In *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, pages 303–304. IEEE.

Santoro, G., Fiano, F., Bertoldi, B., and Ciampi, F. (2018). Big data for business management in the retail industry. *Management Decision*.

Seele, P. (2017). Predictive sustainability control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management. *Journal of Cleaner Production*, 153(C):673–686.

Sharma, K. (2017). *Overview of Industrial Process Automation*. Elsevier Science.

Sivarajah, U., Kamal, M. M., Irani, Z., and Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70:263 – 286.

Stoet, G. and Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological science*, 29(4):581–593.

Takurta, D., Ericson, G., Martens, J., and Severtson, B. (2017). Deployment stage of the team data science process lifecycle - azure.

Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., and Rojo-Álvarez, J. L. (2018). Using big data from customer relationship management information systems to determine the client profile in the hotel sector. *Tourism Management*, 68:187–197.

Van Der Aalst, W. (2016). Data science in action. In *Process Mining*, pages 3–23. Springer.

Xu, L. D., He, W., and Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4):2233–2243.

Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., Zhang, Q., and Yang, L. (2015). Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems*, 17(3):620–630.

Zhou, K., Fu, C., and Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56:215–225.