International Journal of Innovations & Research Analysis (IJIRA)
ISSN :2583-0295, Impact Factor: 5.449, Volume 03, No. 01(II), January- March, 2023, pp 115-118

REVIEW ON ROLE OF DATA SCIENCE IN LIBRARY AND INFORMATION SCIENCE

Rajesh Achra*

ABSTRACT

In order to transform data into information, libraries must adopt new service models. To take advantage of the growing domain of library data analytics, which provides new insights into existing service models, libraries need to improve their technological literacy, especially coding and mark-up. However, acknowledging the importance of proactively calibrating one's professional mission does not imply libraries generate, or manage, big data in the traditional sense. Essentially, it acknowledges that libraries must be aware of being "data savvy" due to the data-intensive world in which they operate. Libraries are not just consumers of data anymore, but also suppliers and to a lesser extent producers. To build a good framework of reference and knowledgebase, patrons rely on good-quality information just as much as new technologies such as artificial intelligence (AI) rely on good-quality data. When determining a reliable data source, data integrity and data ethics are crucial factors to consider. Libraries have always placed a high value on privacy, ethics, and equal access to information, which enables them to serve as new partners for researchers and sources of high-quality data.

Keywords: *Data Science, Library Science, Information Science.*

Introduction

To gain a better understanding of current trends and support more effective decision-making, data science uses statistical and programming methods to extract knowledge from large amounts of data. Our society has produced a tremendous amount of data over the years, so managing and analyzing data sets has become increasingly important. Experts have predicted that there will be an ever-increasing need for data science experts across diverse fields ranging from scientific research and government policies, to commercial marketing. This course of study is tailored towards providing students with the necessary skillset, enabling them to become successful data analysts, data engineers or even data journalists. Furthermore, those willing to further their education could consider taking courses from this program combined with other related programs such as Cultural Heritage Information Management and Information Organization, in order to pursue paths such as data managers, librarians, archivists and curators of research data. [1]

Interdisciplinary skills such as statistics, systems analysis, programming, and specific subject areas are required in data science. Data/text mining, big data analytics, statistics, data visualization, coding, and machine learning are all important skills in the expected skill sets. The goal of data science is to uncover actionable insights hidden in a company's data using math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning, along with specific subject matter expertise. Decision making and strategic planning can be guided by these insights. [1]

With the increasing number of data sources and data, data science has become one of the fastest growing fields across all industries as a result. Thus, it is not surprising that Harvard Business Review named the data scientist the “sexiest job of the 21st century” (link outside of IBM). In order to improve business outcomes, organizations increasingly rely on them to interpret data and make actionable recommendations. Analysts are able to glean actionable insights from data science by using a variety of roles, tools, and processes. [2]

* Master in Library and Information Science, Junior Judicial Assistant, Rajasthan High Court, Jodhpur, Rajasthan, India.

Using a variety of methods, data is collected from all relevant sources, both structured and unstructured. The data can be gathered manually, scraped from the web, or streamed from systems and devices in real time. Besides structured data, such as customer information, unstructured data such as log files, video, audio, pictures, the Internet of Things (IoT), and social media can also be gathered. [2]

- Data storage and data processing is essential for companies to manage their data in various formats and structures. Data management teams develop standards around data storage and structure to enable analytics, machine learning and deep learning models within the company's workflow. Moreover, data preparation procedures such as cleaning, deduplicating, transforming, and combining with ETL jobs or through other integration technologies are necessary for preserving effective data quality before loading into a warehouse, lake, or other repository. [3]
- Data analysis entails an exploratory study to recognize any bias, patterns, ranges and distributions within the data. This assessment serves as a trigger for A/B tests, along with deciding the utility of the data in training models for predictive analytics, machine learning and/or deep learning. If the model produces satisfactory accuracy rates, organizations are on track to make informed decisions that promote scalability. [3]
- Communicate: Finally, insights are presented as reports and other data visualizations so that business analysts and other decision-makers can better comprehend them. It is possible to generate visualizations using a data science programming language such as R or Python. Data scientists can also use specialized visualization tools. [3]

Library and Data Science

The word "data" in data science seems to be a step down from information in libraries, which are known as keepers of information. As an example, big data raw data has no inherent value in its unanalyzed form; on the other hand, library books offer readily accessible insights. [4]

Library science is often referred to as library and information science, which may leave out the concept of data. However, it is important to understand that the raw data used in big data and data science approaches is essential for knowledge production and actionable insight. Therefore, librarians have an essential role to play in leveraging the potential of this area. Presently, people who can analyse raw data and draw knowledge from it are scarce in number. This is where librarians can make a difference and be instrumental in the field of data science. Librarians' expertise with data management and organisation can provide the foundation for training a new generation of data scientists.[4]

In order to mitigate the data scientist shortage, librarians can offer big data services as another tool in the research toolbox. For big data analysis, librarians can play a crucial role in the discovery, understanding, and cleaning of data. In order to be able to analyze data, data must be selected, cleaned, and formatted in a certain way even before the analysis process begins. In the same way that librarians offer resources on many other research topics, they can also offer resources on how to handle big data for analysis. [4]

Librarians offer a range of resources for finding novel data sources, providing background information on subjects, and offering advice on metadata. With Hal Varian pointing out that the capacity to work with data is a valuable ability for the next two decades, it is important that this skill be developed in both data scientists and the general workforce to tackle the scarcity of such talent. This should come as a benefit not only to library patrons, but also librarians, who can use this kind of knowledge in their own careers. As big data becomes more ingrained in today's world, having this know-how is becoming increasingly crucial. [5]

Librarians have been curating knowledge for centuries, making them experts in data access. In a way, they have handled big data before it was called such. As researchers get more involved in this field, library collection development should focus on the techniques behind analytics approaches. The National Institutes of Health's National Library of Medicine provides helpful advice for libraries to present books and resources in areas like machine learning, data management, wrangling and visualization. Additionally, it may be wise to include links to educational websites along with any resources created. [5]

As a matter of fact, libraries are becoming increasingly aware of the potential of data science. With the "Data Scientist Training for Librarians" course, Harvard Library and Harvard-Smithsonian Center for Astrophysics John G. Wolbach Library have teamed up to prepare librarians to handle the growing data needs of their communities. This course gives librarians an overview of the data lifecycle by showing them how to "extract, wrangle, store, analyze, and visualize data." [5]

The Coalition for Networked Information, the National Library of Medicine, and other organizations offer courses that help librarians improve their knowledge and skills. These courses cover topics such as data exploration and analysis, data visualization, data cleaning and preparation, web scraping, bibliometric network analysis, Big Data in Healthcare: Exploring Emerging Roles and more. Through these opportunities, librarians are able to arm themselves with the practices needed to determine the best ways to structure and analyze datasets. Better understanding these methods allows librarians to take on an active role in teaching data science insights through workshops and classes. In doing so they can make a real difference in how Big Data is used in healthcare, business and other sectors. [6]

As mentioned above, librarians are great at organizing and managing information. This skill is also crucial to data science, especially when it comes to curating data for big data applications.[6] It is librarians' skill to communicate strategies and resources that can help researchers and learners learn. As a link between data science and library science, the "new librarian" – a librarian who can offer helpful resources for data scientists – could be so transformative. In order for organizations to make intelligent decisions, data scientists must organize and wrangle large amounts of raw, messy data. [6]

In order to facilitate the creation of new knowledge, library scientists (i.e., librarians) must offer resources. In this case, library science is connected to data science because librarians are able to deal with knowledge and help library patrons find resources to deal with new knowledge so they can gain actionable insights. [7] "The new librarian" who is trained in data science approaches does not need to be a programmer, according to Jeffrey Stanton of Syracuse University iSchool: "Librarians who are interested in knowledge creation should be familiar with how various software tools can transform data, but they do not need to become programmers." It is not necessary for librarians to be database engineers, but they must know how information retrieval tools work. [7] A librarian does not need to be a statistician or graphic designer, but they need a strong grasp of descriptive summaries and basic tests of numerical data, as well as the qualities of successful data displays. Ultimately, it is through the combination of understanding user needs and data curation that librarians can properly carry out their mission. [7]

Librarians have been renowned as promoters of free access to knowledge, strengthening their communities and enlivening collective understanding while safeguarding historical information for future generations. Now, librarians' remits also include more progressive responsibilities than just shelf-stacking: as technology and big data progress, it will become increasingly frequent for people to explore vast datasets to answer questions not addressable through traditional methods such as polls and reviews of extant literature. That's why a new librarian is so important to usher in a new generation of data-driven insights that data scientists and others interested in big data can use. [8]

Data science and big data analytics offer many possibilities to libraries. Not only can patrons benefit from these solutions, but the library itself can take advantage of the technology to make smarter decisions. For instance, librarians can look to checkout records and other information gathered through data analysis to determine what books should be added to their collections. Ultimately, this could help streamline the library's decision-making process for collecting materials for the public. [8]

Data Science Workshops for Library Science

As DS/OS needs vary widely across institutions, disciplines, and researchers, libraries have ample opportunities to provide DS/OS services. As part of the workshop, participants shared DS/OS initiatives and services being offered by their libraries. Each of these services is informed by the participants' understanding of their institutions' unique needs. Based on assessments, surveys, user feedback, and other data inputs, workshop participants leverage their existing skillset as well as secure assistance from partners from inside or outside of the institution. For instance, some participants lacking programming experience sought help from within their institutions or obtained volunteers from The Carpentries - a non-profit organization focused on providing educational opportunities for researchers in coding and computational topics. Currently offered services and infrastructure include: [9]

- "Data science core" - a physical space within the library where patrons can access computers with data-specific tools like R, Python, SPSS, and SAS, as well as staff available to help them.
- Workshops on reproducibility that cover experimental design, registering reports, different peer review formats, fostering an open lab culture, and sharing data, code, and protocols.
- Including best practices for formatting, content and user management, mapping research workflows, features and functionality, and an exploration of different notebook tools.

- Training modules on data literacy and awareness focused on data management principles, data management plans, open science, research rigor, and reproducibility.
- Librarians can participate in lab meetings and work on specific projects within a lab with embedded data services. [9]
- To increase discoverability and encourage data sharing, a data discovery index links multiple institutions.
- Services related to the adoption and use of standards, data quality measures, data sharing processes, data compliance, and data and software curation.
- A librarian performs bibliometrics and research impact assessments, including database searching, data wrangling, data analysis, and data visualization.
- In addition to meeting a range of DS/OS needs, these services and infrastructure fill important institutional gaps for participants. In order to lay the foundation for the discussion that formed the heart of the workshop, participants had to understand how they positioned themselves within the landscape of DS/OS at their institutions. [10]

Conclusion

In spite of its complexity, the transition to the data-science-friendly library is a natural result of recent technological advancements. A librarian who has expertise in knowledge organization and management, as well as the ability to explain how to use and organize information sources, may be the best candidate to help address the shortage of data scientists. As data science continues to be embedded in all industries, citizens need to ensure they can handle the raw data and understand how data science works. Librarians are perfectly positioned to aid this process by ensuring information is stored securely and promoting learning opportunities in the field. Additionally, libraries themselves can benefit from this technology for decisions such as selection, purchasing, preservation and disposal of library resources.

References

1. Virkus, Sirje & Garoufallou, Emmanouel. (2019). Data science from a library and information science perspective. Data Technologies and Applications. ahead-of-print. 10.1108/DTA-05-2019-0076.
2. Surkis, Alisa, Fred Willie ZametkinLaPolla, Nicole Contaxis, and Kevin B Read. “Data Day to Day: Building a Community of Expertise to Address Data Skills Gaps in an Academic Medical Center.” Journal of the Medical Library Association: JMLA 105, no. 2 (April 2017): 185–91. https://doi.org/10.5195/jmla.2017.35.
3. Song, Il-Yeol& Zhu, Yongjun. (2017). Big Data and Data Science: Opportunities and Challenges of iSchools. Journal of Data and Information Science. 2. 2017-2018. 10.1515/jdis-2017-0011.
4. Lyon, Liz &Mattern, Eleanor. (2016). Education for Real-World Data Science Roles (Part 2): A Translational Approach to Curriculum Development. International Journal of Digital Curation. 11. 10.2218/ijdc.v11i2.417.
5. Boston University Libraries. [2016]. *Research data management*. [Online]. Available from:
6. http://www.bu.edu/datamanagement/background/whatisdata/.
7. Calzada Prado J and Marzal MA (2013) Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri: International Journal of Libraries & Information Services* 63(2): 123 – 134.
8. DCC, 2016c. *Glossary.* [Online]. Available from: http://www.dcc.ac.uk/digital-curation/glossary.
9. EPSRC. 2016. *Scope and benefits*. [Online]. Available from:
10. https://www.epsrc.ac.uk/about/standards/researchdata/scope/ .
11. Ingram, C. 2016. *How and why you should manage your research data: a guide for researchers*. [Online]. Available from: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data .
12. University of Edinburgh. 2016a. *A guide to the research data service*. [Online]. Available from:
13. http://www.ed.ac.uk/files/atoms/files/rds_booklet_may2016.pdf..

□❍□