

# Collaboration between Data Scientists and Operations Teams

Ankur Mahida

Subject Matter Expert (SME), Barclays

**Abstract:** *The collaboration between data scientists and operations teams is significant in revolutionizing data - driven research. While the data scientists develop the techniques in research, the other operation teams are equally important in the entire research process by developing the tools for analysis and in the collection of data. This review paper analyzes the gaps in data - driven research, how collaboration between the parties is significant in data research and the solutions for these gaps observed. Additionally, the paper aims to provide insight into the impact and application of collaborative data research especially in firms and the scope involved. Bringing together data scientists through effective communication and language understanding will be effective in the development of new techniques, machine learning, and enhanced algorithms.*

**Keywords:** data science, scientific research, collaboration, machine learning, and operation teams

## 1. Introduction

Conventionally, data science has become a prevalent and integral element in many organizations consisting of data collection, organization, processing, and analysis. Data scientists and researchers can extract data from information sources and construct scientific models for utilizing the information extracted [1]. However, over the years, data scientists haven't been producing the desired results. This is attributed to by the fact that the data scientists may be working in isolation without sufficient collaboration from the operation teams [2]. Most often, operation teams do not participate fully in the data science research processes due to inadequate expertise in the field. Therefore, for effective use of information extracted from data, every party in the research process must play their role to close the gaps in the data science ecosystem.

## 2. Problem Statement

Presently, the collaboration between data scientists and operation teams in all fields has been a controversial topic for discussion. While there might be collaboration between the parties involved, the collaboration is too limited to contribute to the effective collection, storage, and analysis of scientific data [2]. Data science expertise has relied on fragmented teams for analytics leading to gaps in the effective implementation of data science in organizations. Finding the right expertise, data, and tools, thus has become more difficult leading to scientific challenges in data research and innovation. Additionally, there has been rather underpowered research and discoveries in data science even with the growing rates of scientific paper production. Most of these scientific papers cannot reproduce accurate information due to biased findings, especially due to insufficient analysis of data. These factors have been due to a lack of transparency and insufficient collaboration between data specialists. Most importantly, the slow rate of research in data science comes from the gap between data scientists and operation teams (domain scientists, tool developers) [3]. While data scientists may have the required sufficient amount of knowledge of the field of data science, the domain scientists and tool developers often possess limited knowledge in the field not knowing the

appropriate techniques and relying on literature sources. This is attributed to the fragmented data science ecosystem resulting in the overall misunderstanding among the domain scientists, data scientists, and tool developers. As a result, the lack of effective collaboration in data science has often limited innovation and reproducibility in data. Moreover, many innovative researchers in the data science industry may need data science expertise but lack it due to the gap between academia and industry. Similarly, the access to tools and data is limited to data scientists causing the data scientists to work in isolation and limiting their scope of improving scientific research.

## 3. Anticipated Solutions and Enhanced Collaboration

In order to improve the innovations and research in the data science industry, several strategies can be implemented towards collaboration between data scientists and the operation teams. The most vital components towards an effective collaboration in data science will be automating data science, collaboration through team building and cross - functional training, and even enabling easy access to data sources.

Through practices like team building and bidirectional communication between data scientists, domain scientists, and tool developers, collaboration in data science can be enhanced [4]. Feedback on data constraints and needs is equally important in building the relationship in generating scientific research. To enhance the accuracy and reproducibility of the generated results, the generation process must be interlinked between the domain and data scientists, thereby resulting in widespread collaboration and social networking between the domains. Furthermore, the sharing of scientific data between the parties should be instant to foster full collaboration. In short cross - functional training and inter - domain collaboration would help build mutual understanding between data scientists and operation teams.

#### 4. Data Science Automation and Uses

Automation in data science would help mitigate the aftermaths of under-collaboration in data science. Machine learning incorporation in data science would help identify which techniques are required for specific types of data, leading to intelligently optimized algorithms that help save time [5]. This allows for better data analysis and more collaboration between the parties involved, especially with large amounts of data.

Other solutions include organized access to data, and model interpretability which are crucial in closing the gaps and encouraging mutual understanding in data science.

With effective collaboration in data science, there are several anticipated applications, especially in organizations. Insights into data science and analytics are likely to improve inventory management, logistics, and supply chains. Moreover, the interaction between data scientists and domain teams will expand scientific research into machine learning and the development of algorithms in data science [6]. With the revolutionary nature of data science, access to tools and data should be prioritized to realize the true potential of technical algorithms and machine learning. For instance, the OpenML has enabled effective interaction of data scientists as they can interact through data sets, tasks, and flows. In a nutshell, an improved data science domain would result in optimized processes within organizations and communities.

#### 5. Collaboration Impact and Scope

Greater collaboration between data scientists and operations teams has improved business criteria and analytics. This has resulted in a competitive advantage of one organization outperforming another in terms of innovation and productivity [6]. Moreover, advancements in machine learning and algorithms will be greatly impacted by collaboration and enhancement in data science. However, overcoming cultural and communication challenges will surely enable data science to have a great impact on society. Research has shown that firms that employ and use collaborative data science have realized a great improvement in business analytics and success.

While this review paper provides a discussion of the collaboration systems between data scientists and operation teams, providing evidence on greater operations with corresponding collaboration, more research is needed to improve data science. Collaboration models should shift towards internal and external data analytics and emphasize on diverse collaboration strategies [7]. With automated data science and/or integrated machine learning in data science, many organizations will benefit from data science analytics. Consequently, this will provide greater evidence-based guidance for leaders in facilitating partnerships between data scientists and business units.

#### 6. Conclusion

Data science is becoming dominant in most industries prompting the need for research into the accuracy and efficacy

of data results obtained with this scientific research. Making breakthroughs by adopting machine learning, algorithms, and online tools will help revolutionize firms further. Moreover, with the fragmentation of data science among parties; data scientists, and domain scientists, bringing them together will lead to a social network where the rate of discovery and innovation in data-driven science is elevated. This will further facilitate online scientific collaboration, an easy network search engine, transfer of data science information, and automate data science aspects. Further research into building the bridge between data scientists and operation teams of domain scientists and tool developers, especially through cross-functional communication.

#### References

- [1] W. a. W. v. d. A. Van Der Aalst, *Data science in action*, Springer Berlin Heidelberg, 2016.
- [2] S. a. S. J. J. Passi, "Trust in data science: Collaboration, translation, and accountability in corporate data science projects, *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW 1 - 28., 2018.
- [3] J. R. a. R. B. Turner, Collaborative research: Techniques for conducting collaborative research from the science of team science (SciTS), *Advances in Developing Human Resources* 22, no.1: 72 - 86., 2020.
- [4] L. M. a. H. G. Bennett, "Collaboration and team science, " *from theory to practice*, pp.768 - 775, 2012.
- [5] D. J. A. J. D. W. E. O. a. C. D. Wang, "arxiv. org, " 2021. [Online]. Available: <https://arxiv.org/abs/2101.05273>.
- [6] M. A. a. S. E. F. Waller, "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management, " *Journal of Business Logistics* 34, no.2, pp.77 - 84, 2013.
- [7] M. J. R. W. a. K. V. Smith, "FeatureHub: Towards collaborative data science, " *In 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, pp.590 - 600, 2017.

-