International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

# Effective Communication of Data Science Results to Non - Technical Stakeholders

**Sowmya Ramesh Kumar**

Seattle, WA
Email: *rsowmyash[at]gmail.com*

**Abstract:** ***Effective communication of data science results to non - technical stakeholders is pivotal for ensuring the comprehension, acceptance, and practical utilization of data - driven insights within organizations. In a data - centric era where businesses heavily rely on data science for decision - making, bridging the gap between data scientists and non - technical decision - makers is crucial. This involves recognizing the diverse audience, tailoring communication, utilizing visualizations, providing contextual narratives, and incorporating real - world examples. Collaboration and communication within data science workflows are essential, as recent research highlights. The evolving perspective challenges the traditional notion of communication as a final step, emphasizing its continuous relevance throughout the data analysis process. The challenges in communicating machine learning and deep learning model results to non - technical stakeholders are multifaceted. The complexity of these models, their interpretability as "black boxes, " the probabilistic nature of predictions, and the need to convey business impact present significant hurdles. Addressing these challenges requires simplifying concepts, providing intuitive visualizations, and emphasizing practical significance. Regular engagement, training, and fostering a data - driven culture within organizations are vital for enhancing understanding and facilitating informed decision - making among non - technical stakeholders.***

**Keywords:** communication, interpretability, data science, visualization, machine learning, deep learning, collaboration, complexity, simplicity, stakeholders

## 1. Introduction

Effective communication of data science results to non - technical stakeholders is a crucial aspect of ensuring that data - driven insights are understood, accepted, and effectively utilized within an organization. In today's data - driven world, where businesses rely on data science for decision - making, it is imperative to bridge the gap between data scientists and non - technical stakeholders to facilitate informed and strategic decisions.

The first step in effective communication is to recognize the diverse audience involved. Non - technical stakeholders may include executives, marketing professionals, sales teams, or other decision - makers who might not have a deep understanding of the technical intricacies of data science. Tailoring communication to the specific needs and knowledge level of the audience is essential. This involves avoiding jargon and technical details that might overwhelm non - technical stakeholders.

Visualizations play a pivotal role in conveying complex data insights in a comprehensible manner. Graphs, charts, and dashboards can distill intricate analyses into clear and visually appealing representations. Choosing the right type of visualization for the data at hand is crucial; for example, bar charts for comparisons, line charts for trends, and pie charts for proportions. Interactive dashboards also empower non - technical stakeholders to explore data on their own terms.

Context is key in effective communication. Providing a narrative that contextualizes the data, explains its relevance to business objectives, and outlines potential implications is essential. Instead of bombarding stakeholders with raw data, data scientists should craft a story around the findings, making it relatable and applicable to the organization's goals.

In addition to storytelling, incorporating real - world examples and analogies can help bridge the gap between technical and non - technical worlds. Relating data insights to familiar scenarios or industry benchmarks can enhance understanding and make the information more accessible.

## 2. Collaboration and Communication in Data Science

Recent research acknowledges the collaborative nature [Fig 1] of data science and explores strategies to support stakeholders in this collaborative process. Zhang et al. [9] conducted an extensive survey at IBM, identifying five major roles and six key stages in collaborative data science workflows. Communication plays a pivotal role in collaboration, involving both documentation and dissemination processes. Documentation encompasses recording and describing data science work, facilitated by tools like code gathering and provenance collection systems. On the other hand, dissemination involves conveying insights through presentations, reports, and interactive visualizations, with various enterprises and research tools supporting this aspect. While traditional views often treat communication as a final step, recent studies, such as Voder's interactive widgets and NB2Slides for computational notebooks, recognize the importance of intermediate communication throughout the data analysis process. This evolving perspective challenges the notion that communication solely occurs in the final stages of collaborative data science workflows, emphasizing its continuous relevance during analysis.

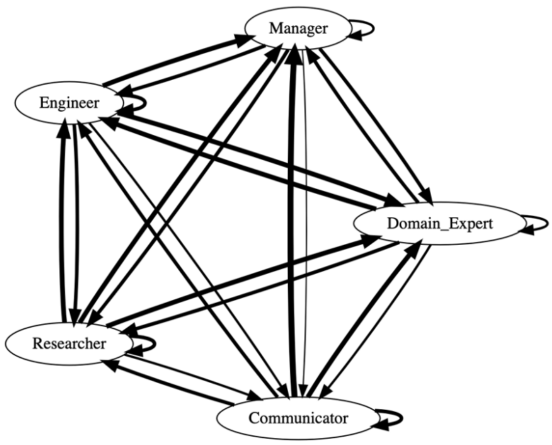


**Figure 1:** Collaboration among roles in data science projects [1]

Data science is rapidly advancing and gaining widespread adoption, but there is a notable absence of well - established methods to quantify and communicate uncertainty associated with information in large datasets. While uncertainty communication has been explored in geospatial, visualization, and cognition studies, there is still a lack of effective and broadly applicable approaches. Decision science has extensively examined decision - making under uncertainty, and addressing uncertainty in data science is challenging due to the relatively new state of the field. However, this also presents an opportunity to influence the emerging field by incorporating relevant disciplines from the beginning.

Educational programs in data science are emerging, and the growing demand for trained professionals in this field suggests an increase in their numbers. Therefore, associated fields like computer science, statistics, and decision science can influence curricula and pedagogy by incorporating training in analytical and behavioral aspects of error and uncertainty.

For example, simulation science primarily deals with large - scale computation models, falling under numerical analysis, scientific or high - performance computing, and application - specific engineering. In contrast, data science, focusing on the analysis of existing data, draws expertise from AI, machine learning, and data mining. Although simulation science and data science play similar roles in the data - to - decision pipeline, they differ in problem framing, computational tools, and result communication mechanisms. While uncertainty is intrinsic to statistics, closer collaboration is needed to address fundamental problems and technologies for quantifying uncertainty in both data and simulation science.

Various academic disciplines study the communication of uncertainty to decision - makers and how they utilize such information. Scientific visualization primarily focuses on pictorial representations of uncertainty in high - dimensional outputs of physical simulations. The information visualization community has addressed this to a lesser extent, reflecting a similar situation in data science. Despite vast research literature and applications in fields like health, environment, and finance, cognitive and decision sciences have had limited connections with computer science, except for essential work on human - computer interaction.

Fig 2 highlights an important issue in addressing technical challenges associated with uncertainty in computation, which is that the relevant research is fragmented and does not map well onto the individual steps in the data - to - decision pipeline. These current interactions of different fields with the work flow are depicted as grey ovals in the figure.

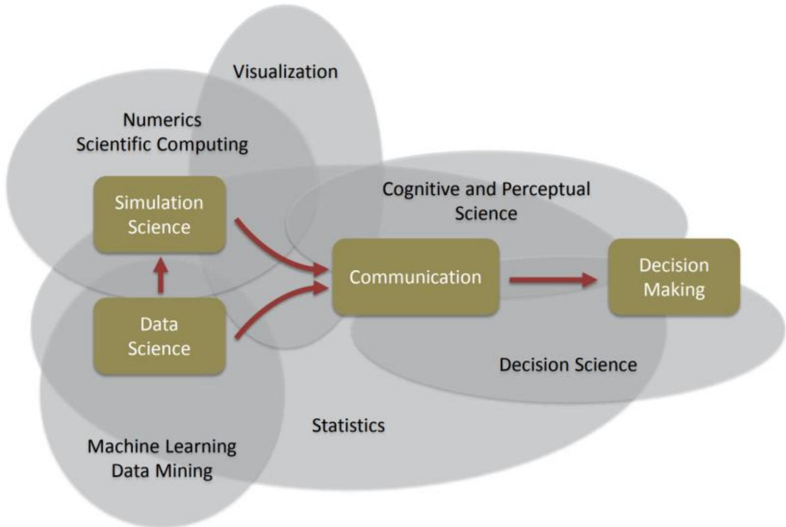


**Figure 2:** Academic disciplines relevant to understanding the data - to - decision pipeline [8]

## 3. Challenges of communicating Machine Learning / Deep learning models

Communicating machine learning (ML) and deep learning (DL) model results to non - technical stakeholders pose several challenges. One major hurdle is the inherent complexity of these models. ML and DL involve intricate algorithms and layers of neural networks, making it challenging to convey the intricacies without overwhelming non - technical audiences.

Another challenge lies in the interpretability of these models. Unlike traditional statistical methods, ML and DL models often function as "black boxes, " making it difficult to explain how specific decisions are reached. This lack of transparency can lead to skepticism and hesitation among non - technical stakeholders who may find it challenging to trust and act upon results they cannot comprehend.

Additionally, addressing the uncertainty associated with model predictions is crucial. Non - technical stakeholders might not be familiar with the probabilistic nature of ML models and may struggle to grasp confidence intervals or prediction intervals. This lack of understanding can hinder their ability to make well - informed decisions based on the model outputs.

Furthermore, conveying the business impact of ML and DL results is essential. Stakeholders may prioritize actionable insights rather than technical metrics. Bridging the gap between technical metrics like accuracy or precision and their real - world implications is vital for effective communication. In overcoming these challenges, a concerted effort to simplify complex concepts, provide intuitive visualizations, and emphasize the practical significance of model outcomes is essential. Regular engagement, training, and fostering a data - driven culture within the organization can enhance understanding and facilitate more informed decision - making by non - technical stakeholders.

## 4. Conclusion

Effectively communicating data science results is not only about conveying technical findings but also about fostering collaboration and understanding among diverse stakeholders. The evolving landscape of data science workflows emphasizes the continuous nature of communication throughout the analysis process. Recognizing this, organizations should prioritize strategies that enhance communication skills, leverage visualizations, and provide real - world context.

In the realm of machine learning and deep learning, transparency, interpretability, and the translation of technical metrics into actionable insights are key challenges. Overcoming these challenges requires a concerted effort to demystify complex models, make uncertainty comprehensible, and highlight the tangible business impact. As organizations increasingly embrace data - driven decision - making, empowering non - technical stakeholders with the ability to grasp, trust, and act upon data science results becomes imperative for overall success.

## References

[1] Pang, R. Y., Wang, R., Nelson, J., & Battle, L. (2022, October). How Do Data Science Workers Communicate Intermediate Results? In *2022 IEEE Visualization in Data Science (VDS)* (pp.46 - 54). IEEE.

[2] Zhang, A. X., Muller, M., & Wang, D. (2020). How do data science workers collaborate? roles, workflows, and tools. Proceedings of the ACM on Human - Computer Interaction, 4 (CSCW1), 1 - 23.

[3] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. IEEE transactions on visualization and computer graphics, 25 (1): 22–31, 2018.

[4] A. Batch and N. Elmqvist. The interactive visualization gap in initial exploratory data analysis. IEEE transactions on visualization and computer graphics, 24 (1): 278–287, 2017. [4] A. Bauerle, ¨ A. A. Cabrera, F. Hohman, M. Maher, D. Koski, X. Suau, ´ T. Barik, and D. Moritz. Symphony: Composing interactive interfaces for machine learning. In CHI Conference on Human Factors in Computing Systems, pp.1–14, 2022.

[5] C. L. Borgman, J. C. Wallis, and M. S. Mayernik. Who's got the data? interdependencies in science and technology collaborations. Computer Supported Cooperative Work (CSCW), 21 (6): 485–523, 2012. [6] V. Braun and V. Clarke. Using thematic analysis in psychology. Qualitative research in psychology, 3 (2): 77–101, 2006.

[6] M. Brehmer and R. Kosara. From jam session to recital: Synchronous communication and collaboration around data in organizations. IEEE Transactions on Visualization & Computer Graphics, 28 (01): 1139–1149, 2022.

[7] A. T. Chatfield, V. N. Shlemoon, W. Redublado, and F. Rahman. Data scientists as game changers in big data environments.2014.

[8] Whitaker, R., Thompson, W., Berger, J., Fischhof, B., Goodchild, M., Hegarty, M.,. . . & Wendelberger, J. (2020). Workshop on Quantification, Communication, and Interpretation of Uncertainty in Simulation and Data Science. arXiv preprint arXiv: 2004.12939.

[9] D. Wang, Q. V. Liao, Y. Zhang, U. Khurana, H. Samulowitz, S. Park, M. Muller, and L. Amini. How much automation does a data scientist want? arXiv preprint arXiv: 2101.03970, 2021.