Communities and Data Revolution

FatanehVahabi[1,2*]

[1]*Iranian Research Institute for Information Science and Technology, Tehran, Iran*

[2]*Isfahan University of Medical Sciences, Isfahan, Iran*

***Corresponding Author**
FatanehVahabi, Iranian Research Institute for Information Science and Technology, Tehran, Iran.

Abstract
The data revolution is transforming the way we interact with the world. With the rise of big data, organizations have access to an unprecedented amount of data, enabling them to better inform decision-making and develop new products and services. However, to capitalize on this data revolution, it is essential to build strong, connected communities to make full use of the data.By engaging stakeholders and developing a diverse and engaged community, organizations can better understand how to use data to gain insights, create new opportunities, and solve problems. This includes engaging with citizens, businesses, civil society organizations, and local government to ensure a wide range of perspectives are heard and incorporated into the data strategy. This kind of engagement also provides a platform for public dialogue and building trust in the data-driven decision-making process.Organizations should also invest in data literacy initiatives to ensure that data is used effectively. These initiatives should be tailored to the needs of the community and focus on educating stakeholders on the responsible use of data. This could include workshops, training sessions, and hackathons to familiarize the community with data analysis, data visualization, and data storytelling.Finally, organizations should develop data sharing protocols to ensure that data is shared securely and ethically. This includes having clear terms and conditions regarding data usage, data protection, and data security. It also involves developing policies to ensure that data is used responsibly and in the public interest.

1. Introduction

The Data Revolution has transformed the way we collect, analyze, and utilize information. This rapid evolution, spurred by advancements in technology, has created vast opportunities and challenges. In order to navigate this complex landscape, building communities that foster collaboration and knowledge sharing has become crucial. These communities play a vital role in harnessing the immense potential of the Data Revolution and maximizing its impact. By creating a data-driven community, organizations can harness the data revolution to create meaningful change. This could include developing innovative products and services, improving public services, and creating economic opportunities. By investing in building strong, connected communities, organizations will be better placed to unlock the potential of the data revolution. The digital age has created a new era of data revolution, one that has the potential to transform how we interact with the world. By leveraging the power of technology, communities can build upon the data revolution to develop meaningful relationships, foster creativity, and create a better understanding of the world around us. The digital age has provided an unprecedented level of data, which can be used to better understand the world around us. This data can be used to develop a better understanding of the needs of a community, to identify problems and opportunities, and to create solutions. By harnessing the data revolution, communities can create a more connected and efficient society. Data can be used to improve the access to information and resources within a community. Through data-driven decision making, communities can better understand the needs of their citizens, and develop solutions that meet those needs. Data can also be used to identify problems and opportunities within a community, providing valuable insight into how to create an environment that encourages growth and development.

In addition to using data to identify problems and opportunities, communities can also use data to create meaningful relationships. By leveraging the power of technology, communities can create social networks that allow people to connect, share ideas, and collaborate. Technology can also be used to foster creativity, by providing a platform for people to develop and share new ideas. The data revolution can also be used to create a better understanding of the world around us. By aggregating data from multiple sources, communities can generate a more accurate picture of the world, and use that data to create solutions that are tailored to the needs of the community. In addition, data can be used to inform policy decisions, allowing for greater transparency and accountability. The data revolution has the potential to transform how communities

interact with the world. By leveraging the power of technology and data, communities can build upon the data revolution to develop meaningful relationships, foster creativity, and create a better understanding of the world around us.

- Identify potential stakeholders: Identify a range of potential stakeholders, including individuals, organizations, and government agencies, who could contribute to the effort.
- Develop a strategy: Develop a strategy to engage these stakeholders and create a forum for collaboration.
- Establish a governing body: Create a governing body to manage the efforts and ensure that stakeholders are working together effectively.
- Design a data governance framework: Design a data governance framework to ensure the data is handled safely and ethically.
- Identify data sources and use cases: Identify potential data sources and use cases.
- Create a data access and sharing plan: Develop a plan for sharing data securely and ethically.
- Develop data literacy resources: Develop resources such as tutorials and workshops to help people understand the power of data and how to use it responsibly.
- Advocate for the use of data: Advocate for the use of data to improve decision-making and optimize performance.
- Monitor and evaluate progress: Monitor and evaluate progress to ensure the strategy is successful.

1.1 Understanding the Data Revolution
The Data Revolution refers to the accelerated growth and impact of data in various domains of society. It encompasses the increasing availability and accessibility of data, along with the rise of data-driven decision-making. Today, data is generated at an astonishing rate, with every online transaction, social media post, and sensor recording contributing to this vast pool of information. This abundance of data provides valuable insights into various aspects of our lives, such as healthcare, transportation, and climate change.

1.2 The Role of Communities in the Data Revolution
Communities form the backbone of the Data Revolution. Within these communities, individuals, organizations, and institutions collaborate, share knowledge, and work towards common goals. By pooling their collective expertise and resources, these communities stimulate innovation, problem-solving, and the development of best practices. This collaborative environment empowers individuals and organizations to make informed decisions, address challenges, and leverage the potential of data to its fullest extent.

1.3 Benefits of Building Communities in the Data Revolution
Building communities within the Data Revolution offers numerous benefits. Firstly, it empowers individuals and organizations by providing them with the tools, knowledge, and support necessary to effectively utilize data. By sharing their experiences, challenges, and successes, community members can learn from one another and enhance their data literacy.

Secondly, building communities fosters interdisciplinary collaboration. The complex nature of data requires expertise from various fields, including computer science, statistics, social sciences, and ethics. By bringing together individuals from diverse backgrounds, communities facilitate the exchange of ideas and the creation of innovative solutions. This interdisciplinary approach not only enhances the quality of analysis but also ensures a broad perspective when addressing complex societal challenges.Lastly, building communities helps address ethical and privacy concerns associated with the Data Revolution. By promoting discussions and setting standards within the community, ethical guidelines for the responsible use of data can be developed. This contributes to a more transparent and accountable data ecosystem, where individuals' rights and privacy are respected.

1.4 Challenges and Solutions
Despite the numerous benefits, building communities in the Data Revolution also faces challenges. One such challenge is the existence of data silos, where data is trapped within specific organizations or sectors without being effectively shared or utilized. To overcome this, communities can promote data openness and encourage collaboration, thus breaking down existing silos and fostering a more integrated approach to data management.

Another challenge is ensuring inclusivity and diversity within data communities. Often, certain demographic groups, such as women and marginalized communities, are underrepresented in the field of data science. By actively promoting inclusivity and diversity, communities can tap into a wider range of perspectives, talents, and experiences. This not only enriches the pool of knowledge but also ensures that the benefits of the Data Revolution are accessible to all.

Finally, institutional support and infrastructure are crucial for the sustainable growth of data communities. Governments, academia, and industry must recognize the importance of building communities and invest in initiatives that support their formation. This includes providing funding, infrastructure, and policies that enable collaboration, knowledge sharing, and innovation.

1.5 Real-world Examples
Several real-world examples demonstrate the power of building communities in the Data Revolution. Open-source data communities, such as the Apache Software Foundation and the Linux Foundation, provide platforms for collaboration and innovation in the development of data tools and technologies. These communities enable diverse individuals and organizations to contribute to and benefit from the collective wisdom of the community.Data for social good initiatives, such as Data Kind and the Data Science for Social Good Fellowship, demonstrate the potential of communities in addressing societal challenges. By bringing together data scientists, social scientists, and domain experts, these initiatives leverage data to drive social change, improve public service delivery, and solve complex social problems. Government-led data initiatives, such as the Open Government Data (OGD) platforms, promote transparency and

citizen engagement by opening up government-held data to the public. These initiatives provide the foundation for communities to form, allowing citizens, researchers, and entrepreneurs to collectively analyze data, develop insights, and contribute to evidence-based decision-making.

2. Conclusion

Building communities is essential to harness the potential of the Data Revolution. These communities' foster collaboration, knowledge sharing, and innovation, empowering individuals and organizations to effectively utilize data. By addressing challenges, such as data silos and inclusivity, and receiving institutional support and infrastructure, communities can maximize the benefits of the Data Revolution. As individuals, we have a responsibility to actively contribute to and benefit from these communities, ensuring that the power of data is harnessed for the betterment of society.

Result

In today's digital era, data has become the driving force behind innovation and progress. The abundance of data presents both opportunities and challenges. To fully harness the potential of the data revolution, it is crucial to build communities that can effectively navigate this vast ocean of information.Building communities to harness the data revolution requires a multi-faceted approach. It involves creating platforms and spaces where individuals and organizations can come together to share insights, collaborate, and create new knowledge. These communities act as catalysts, encouraging the exchange of ideas and best practices. Moreover, building communities also entails fostering a culture of data literacy and education. By providing training and resources, individuals can acquire the skills needed to effectively analyze and interpret data. This empowers them to make informed decisions and drive meaningful change in their respective fields.

Another critical aspect of building communities is promoting inclusivity and diversity. Encouraging participation from a wide range of backgrounds and perspectives ensures a holistic and comprehensive understanding of the data landscape. By incorporating diverse viewpoints, communities can make better-informed decisions and develop innovative solutions.In conclusion, building communities to harness the data revolution is imperative in today's data-driven world. By creating spaces for collaboration, promoting data literacy, and embracing diversity, we can unlock the true transformative power of data and drive sustainable progress for the benefit of all.

Declarations

I give my consent for information about my article to be published in journal of big data.

Data Not Available

The author declare no competing interests.

References

1. Johnson, J. M., &Khoshgoftaar, T. M. (2019). Medicare fraud detection using neural networks. *Journal of Big Data, 6*(1), 63.
2. Matsuzaka, Y., Hosaka, T., Ogaito, A., Yoshinari, K., &Uesawa, Y. (2020). Prediction model of aryl hydrocarbon receptor activation by a novel qsar approach, deepsnap–deep learning. *Molecules, 25*(6), 1317.
3. Johnson, J. M., &Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of big data, 6*(1), 1-54.
4. Spadon, G., Carvalho, A. C. D., Rodrigues-Jr, J. F., & Alves, L. G. (2019). Reconstructing commuters network using machine learning and urban indicators. *Scientific reports, 9*(1), 11801.
5. Anghel, A., Papandreou, N., Parnell, T., De Palma, A., &Pozidis, H. (2018). Benchmarking and optimization of gradient boosting decision tree algorithms. *arXiv preprint arXiv:1809.04559.*
6. Chen, T., &Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
7. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
8. Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., &Bauder, R. A. (2020). Investigating class rarity in big data. *Journal of Big Data, 7,* 1-17.
9. Herland, M., Khoshgoftaar, T. M., &Bauder, R. A. (2018). Big data fraud detection using multiple medicare data sources. *Journal of Big Data, 5*(1), 1-21.
10. Sheshasaayee, A., & Lakshmi, J. V. N. (2017, July). An insight into tree based machine learning techniques for big data analytics using Apache Spark. *In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (pp. 1740-1743). IEEE.
11. Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... &Stoica, I. (2016). Apache spark: a unified engine for big data processing. Communications of the ACM, 59(11), 56-65.
12. 12. Ensembles—rdd-based api. 2020.
13. Hasanin, T., Khoshgoftaar, T. M., &Leevy, J. L. (2019, July). A comparison of performance metrics with severely imbalanced network security big data. *In 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 83-88). IEEE.
14. Sudha, P., &Gunavathi, R. (2016). A survey paper on map reduce in big data. *International Journal of Science and Research, 5*(9).
15. Khramtsov, V., Sergeyev, A., Spiniello, C., Tortora, C., Napolitano, N. R., Agnello, A., ... &Shulga, V. (2019). Kids-squad-ii. machine learning selection of bright extragalactic objects to search for new gravitationally lensed quasars. A56.
16. Daoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering, 13*(1), 6-10.
17. Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting

loan default in peer-to-peer lending using narrative data. *Journal of Forecasting, 39*(2), 260-280.

18. Bayne, L. E. (2018). Big data in neonatal health care: big reach, big reward?. *Critical Care Nursing Clinics, 30*(4), 481-497.

19. Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics, 15*(3), 20170030.

20. Bellazzi, R. (2014). Big data and biomedical informatics: a challenging opportunity. *Yearbook of medical informatics, 23*(01), 08-13.

21. Sinha, A., Hripcsak, G., &Markatou, M. (2009). Large datasets in biomedicine: a discussion of salient analytic issues. *Journal of the American Medical Informatics Association, 16*(6), 759-767.

22. Scruggs, S. B., Watson, K., Su, A. I., Hermjakob, H., Yates III, J. R., Lindsey, M. L., & Ping, P. (2015). *Harnessing the heart of big data. Circulation research, 116*(7), 1115-1119.

23. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications, 19*, 171-209.

24. 24. IEEE Internet Computing. IEEE Internet Computing, 2012: 1–6.

25. 25. Bolouri, H. (2014). Modeling genomic regulatory networks with big data. *Trends in Genetics, 30*(5), 182-191.

26. 26. Dinov, I. D. (2016). Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *Gigascience, 5*(1), s13742-016.

27. Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. *Kidney research and clinical practice, 36*(1), 3.

28. Ossio, R., Roldan-Marin, R., Martinez-Said, H., Adams, D. J., & Robles-Espinoza, C. D. (2017). Melanoma: a global perspective. *Nature Reviews Cancer, 17*(7), 393-394.

29. Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: promise and challenges. *Nature Reviews Cardiology, 13*(6), 350-359.

30. Fernandez-Luque, L., Mejova, Y., Mayer, M. A., Hasvold, P. E., & Joshi, S. (2016). Panel: big data & social media for empowering patients with diabetes. *In Nursing Informatics 2016* (pp. 607-609). IOS Press.

31. Feldman, K., & Chawla, N. V. (2015). Does medical school training relate to practice? evidence from big data. *Big data, 3*(2), 103-113.

32. Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical teacher, 36*(3), 216-222.

33. O'Sullivan, D. E., Brenner, D. R., Demers, P. A., Villeneuve, P. J., Friedenreich, C. M., King, W. D., &ComPARe Study Group. (2019). Indoor tanning and skin cancer in Canada: A meta-analysis and attributable burden estimation. C*ancer epidemiology, 59,* 1-7.

34. Xu, H., Aldrich, M. C., Chen, Q., Liu, H., Peterson, N. B., Dai, Q., ... & Denny, J. C. (2015). Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *Journal of the American Medical Informatics Association, 22*(1), 179-191.

35. Nguyen, D. C., Pathirana, P. N., Ding, M., & Seneviratne, A. (2020). Blockchain for 5G and beyond networks: A state of the art survey. *Journal of Network and Computer Applications, 166*, 102693.

36. Berdik, D., Otoum, S., Schmidt, N., Porter, D., &Jararweh, Y. (2021). A survey on blockchain for information systems management and security. *Information Processing & Management, 58*(1), 102397.

37. Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., &Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications, 71,* 151-166.

38. Abubakar, A. M., Behravesh, E., Rezapouraghdam, H., &Yildiz, S. B. (2019). Applying artificial intelligence technique to predict knowledge hiding behavior. *International Journal of Information Management, 49,* 45-57.

39. Ali, S., Danish, M., Khuwaja, F. M., Sajjad, M. S., & Zahid, H. (2019). The intention to adopt green IT products in Pakistan: Driven by the modified theory of consumption values. *Environments, 6*(5), 53.

40. Young, T., Cambria, E., Chaturvedi, I., Zhou, H., Biswas, S., & Huang, M. (2018, April). Augmenting end-to-end dialogue systems with commonsense knowledge. *In Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

41. Zadeh, A., Liang, P. P., Poria, S., Vij, P., Cambria, E., &Morency, L. P. (2018, April). Multi-attention recurrent network for human communication comprehension. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

42. Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016, March). The limitations of deep learning in adversarial settings. *In 2016 IEEE European symposium on security and privacy (EuroS&P)* (pp. 372-387). IEEE.

43. Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016, May). Distillation as a defense to adversarial perturbations against deep neural networks. *In 2016 IEEE symposium on security and privacy (SP)* (pp. 582-597). IEEE.

44. Wahabi, F., Asadi, S., &Ganjehtar, S. (2021). Neural Network Modelling of the Information Behavior of Database Users Based on their Previous Interactions with the Search Results. *Iranian Journal of Information Processing and Management, 37*(1), 255-276.

45. Vahabi, Fataneh, Asadi, Saeed.2019. Discovering user click patterns for use in database queries. *Science and techniques of information management,6*(4): 137-158.