Topic 4

Research for Data Science

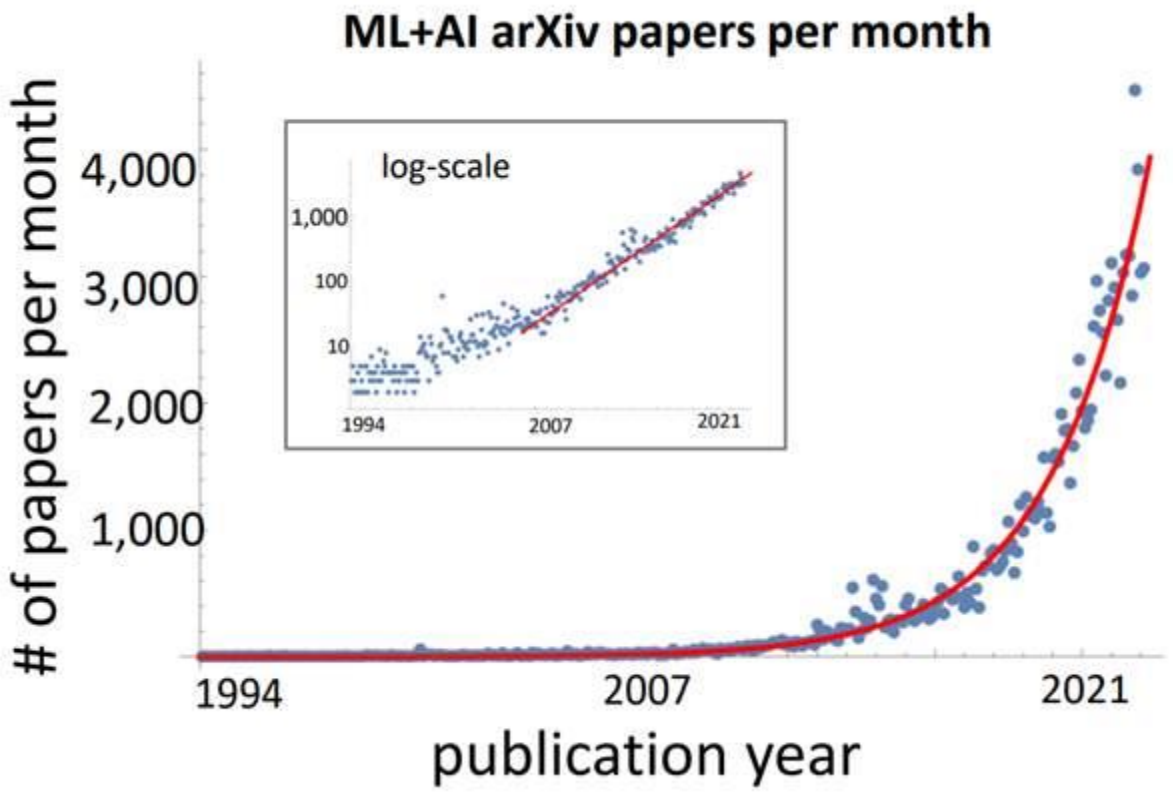
The Research Process

Research is the systematic investigation into a topic to gain understanding and knowledge. In Data Science, data-related topics are researched to gain a deeper understanding of the data and uncover breakthroughs in the field.

Research can benefit data scientists by building a portfolio, as it helps develop more impactful projects, inspire new insights, and gain a deeper perspective of their problems.

Growth of Data Science Research

In recent years, the growth of Artificial Intelligence (AI) and Data Science research has been exponential. As technology advances, researchers are making ever-greater strides in the development of AI and Data Science, allowing us to take on complex tasks that were once unattainable.



Finding Research

Finding research is essential for any data science project. It helps to:

- Develop a deeper understanding as well as a broader perspective on the data
- Identify the latest trends and breakthroughs
- Make informed decisions about your approach
- Provide an opportunity to identify potential challenges and develop strategies for mitigating them

Google Scholar

Google Scholar is a search engine for scholarly literature. It allows users to search for articles, books, theses, court opinions, and other academic content. It also provides links to the full-text versions of the range where available.

Google Scholar provides an easy way to discover and access scholarly content from various sources.

https://scholar.google.com/

ArXiv

ArXiv is a repository of open-access scientific papers in mathematics, physics, astronomy, computer science, quantitative biology, statistics, and finance. It is maintained by Cornell University Library and is hosted on the Cornell University website.

The repository allows users to download, share, and comment on preprints or draft versions of articles before being peer-reviewed and published.

https://arxiv.org/

Papers with Code

Paperswithcode is a platform for researchers to find the latest Machine Learning research papers, associated code, and state-of-the-art performance on various tasks.

It helps researchers stay up-to-date with the latest advances, compare their work to the state-of-the-art developments, and track trends in the field.

https://paperswithcode.com/

Top Data Science Publication Sources

1. NeurIPS (Neural Information Processing Systems)(opens in a new tab)
2. Nature Machine Intelligence(opens in a new tab)
3. International Conference on Learning Representations(opens in a new tab)
4. European Computer Vision Association(opens in a new tab)
5. IEEE Transactions on Knowledge and Data Engineering:(opens in a new tab)
6. IEEE Transactions on Pattern Analysis and Machine Intelligence(opens in a new tab)
7. ACM Transactions on Intelligent Systems and Technology(opens in a new tab)
8. Journal of Machine Learning Research(opens in a new tab)
9. International Journal of Machine Learning and Cybernetics(opens in a new tab)
10. Knowledge-Based Systems:(opens in a new tab)
11. Data Mining and Knowledge Discovery

Data Science article aggregators

Data Science aggregator sites like KDNuggets, Analytics Vidhya, Towards Data Science, and Kaggle Blog are online platforms that curate and collect news, articles, and resources related to data science.

They are helpful for research because they often publish articles on groundbreaking impactful research publications, and they are a great way to keep up to date with the latest innovations in the field.

How to read research papers

Reading research papers can be a daunting task for machine learning practitioners due to the complex language, equations, diagrams, and jargon and the vast amount of information to digest.

Understand the paper's objective - this can be done by reading the abstract and scanning the paper's introduction to get a general sense of the paper's structure.

Read the paper's conclusion – this can help you determine the main points the authors were trying to make and what they achieved.

Review the paper's structure – this can be done by reading the headings, subheadings, and sentences at the beginning and end of each paper. Pay attention to diagrams as they can often help you gain perspective on process flows, results and the dataset used.

Literature Reviews

A literature review is something to consider including in your portfolio if you wish to demonstrate an in-depth knowledge of a particular subject.

Gain an in-depth understanding of a topic

Provide an overview of the current state of knowledge

Identify gaps in knowledge, develop hypotheses, and inform the design of research studies

Compare different studies and draw conclusions about their findings.

Great Literature Reviews

Focus on finding the most relevant research

When conducting a literature review for a data science project, it is important to focus on the most relevant and recent research related to the project. This will help ensure that the project is well-grounded in the field's most up-to-date knowledge and research.

Utilize a variety of sources

Be sure to use a variety of sources when conducting your literature review. This includes primary sources such as journal articles and secondary sources such as books and websites.

Great Literature Reviews

Identify the most important research

As you read the literature, identify the most important research related to the project. This will help you to focus on the most relevant information and show the reader of your portfolio that you are well-versed in the area.

Create a visual representation of your research

One way to make your data science portfolio stand out is to create a visual representation of your literature review. This could be a concept map or infographic, making your research easier to understand and digest.

Great Literature Reviews

Utilise research tools

Using databases and citation tools can help you quickly identify relevant research and save time when conducting your literature review.

Ten Simple Rules for Writing a Literature Review