Upgrade by February 15 and save 15% [$127.50 ~~$150~~]
Use code **EDXWELCOME** at checkout! **Upgrade Now**

Audit Access Expires Feb 25, 2020
You lose all access to this course, including your progress, on Feb 25, 2020.
Upgrade by Feb 29, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

Literature reviews and reliable sources

Undertaking a literature review

One of the most indispensable skills required to navigate the ever-growing field of analytics is learning to digest information on new solutions. In this reading, we will focus on some basic principles that will help you on your way.

Independent thinking and critical analysis

An essential part of academic education is not to force knowledge upon people, but rather to facilitate them in finding their own truths. In light of this, it is important to develop the skill of independent thinking and critical analysis. This entails informing yourself of what is going on around you. Naturally, we resort to the internet these days and with its plethora of search engines it is easier than ever to find information. The key is to be able to digest the value of that information. In this reading, I have given you some tips to guide you when acquiring sources for the various research activities you will complete throughout the course.

Search engines and reliable sources

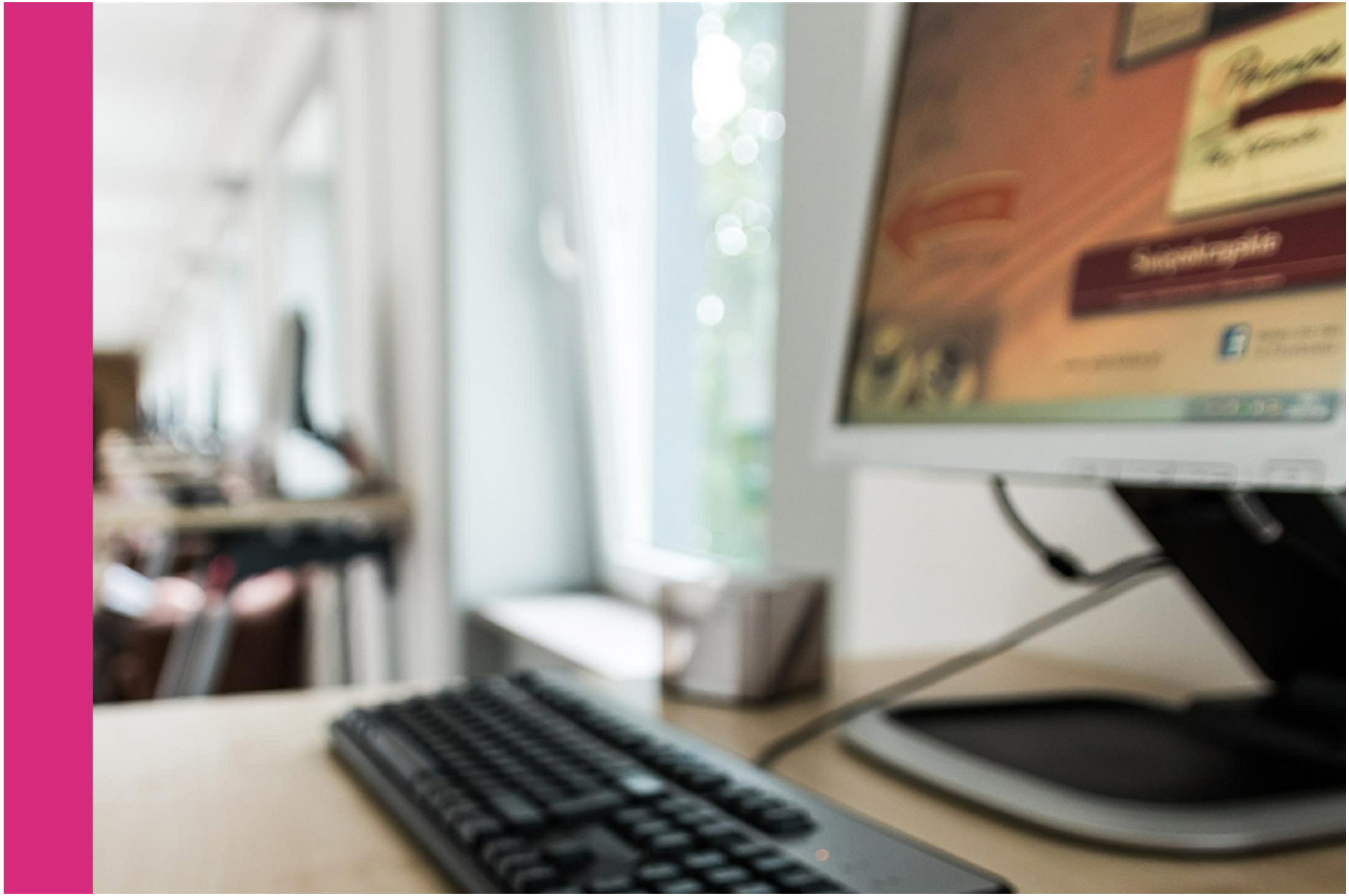
Let's look at a simple example: linear regression. Searching for the term 'linear regression' will yield many hits on a search engine like Google. How will you be able to tell whether a source is a good one? First of all, Google has probably done some work for you already as top hits in analytics are typically quite trustworthy. As a general bit of advice, try to look into the reputation of the source. For example, is it another university or a site with many visitors? The tricky part comes from discussion forums. Naturally, many people have the same questions as you do. You might find an answer but be sure that there is solid validation. The background of the provider of the answer can help, the number of upvotes and so on. As a rule, try to find at least three sources and only consider answers for which there is agreement among different sources.

Your default position should still be to find textbooks online that give you a verified answer.

Various excellent sources for information include:

- Wikipedia, especially when provided with adequate referencing;
- Stackoverflow, mathoverflow, math.stackexchange/stats.stackexchange;
- Course notes from universities; and
- Implementation notes of big software packages (for example, scikit-learn, R, pandas, Weka, and so on).

However, research is not just about finding an answer to a certain question. It is actually about compiling an idea of what is out there. If you are tasked to find the latest developments in linear regression, your scope changes altogether. It is this type of research that is required for the course.



Library (2014) by jarmoluk on Pixabay, licensed under CC0 (modified image)

Academic and subject-specific sources

There are many sources out there, and for predictive analytics, Google Scholar, Scopus, JSTOR, and DBLP form a good basis to find most academic sources. Scopus, JSTOR, and DBLP contain mainly published sources. This published bit is important as it means that it is verified by the community. Google Scholar also contains works-in-progress. This is important, as the field of analytics is growing and expanding so rapidly, that the publishing cycle sometimes has a hard time catching on. Regardless, there are a few heuristics that can help you in selecting published and unpublished material:

- How many works and citations do the authors have. More importantly perhaps, how many do the first two authors (who typically do the majority of the work and writing) have?
- What is the outlet? Is it a journal or conference? If so, are they regarded as quality outlets? While this is very much up for debate within the academic community, one can take into account rankings and metrics such as the impact factor. They can be found easily for journals by visiting their homepage. For conference, one can check rankings. A personal favourite of mine is http://portal.core.edu.au/conf-ranks/ for conferences, and its journal equivalent http://portal.core.edu.au/jnl-ranks/.

Are there authors with no publications and citations that published a paper in unknown journals? Yes, there are plenty. The final filter has to be you. Your intuition might be as good an indicator of the quality of a work as any!

Of course, this is an extremely simplified and coarse way of doing things, but these tips might give you a good entry point to work on a particular subject. Citations and references make every work connected to others. If you find a work in a good journal, typically it has an excellent related work section. That will help you on your way as well.

Helpful tips and practices

A good practice is always to summarise every paper you come across. This will help you capture the information in that work and makes comparisons later easier. As an example, if you were trying to check for the applicability of linear regression for measuring pollution, it might be worthwhile to keep track of the following things in a paper:

- What variables are used in the model?
- How is the pollution measured?

- What scale is used?
- What dataset is used?
- What other techniques are benchmarked?
- What are the key takeaways?

By keeping track of these characteristics, it becomes easy to quickly summarise various papers and compare them. It might be that some papers are contradicting each other, or that some find that particular variables only work for particular datasets. Furthermore, it becomes easy to spot gaps. Perhaps there are variables that haven't been tested for a particular dataset. This might give you a good idea for a research paper or independent research, for example, in a dissertation or capstone project.

Hopefully, this brief overview has given you a few good pointers for doing your own investigation in the future.

Learn About Verified Certificates

© All Rights Reserved