

How to develop yourself with Data Science

Whitepaper

Ronald Buijsse - Academic Director at JADS



Introduction

Being a successful data scientist takes more than just knowing how to make sense out of big sets of data. You also need to understand business issues, decision making processes and how to address users and management with new insights.

Therefore, starting and developing a data science career requires many decisions. The main steps to decide on are:

1. Step 1: What is the right ambition level and type of education to choose
2. Step 2: What technologies to learn
3. Step 3: What project methodology to master
4. Step 4: How to secure the impact of your data science work with a focus on decision making

In the text below we will share our main best practices on these four steps.

Step 1: Choose your ambition level and type of education

Learning Data Science is a popular way to boost careers. Overall, at the start, there are two choices to make: your ambition level and your type of education.

Based on a dataset involving 82 professionals, studying Data Science, we learned that ambition level was the main feature to explain the level of learning and business impact created during the part-time study period of one year. Other features which turned out to have less impact were a.o.: earlier type of education (all participants were on bachelor level or higher), working

environment (already data science related work or not), programming experience, mathematical background, project management experience, business experience and hours spend on online data science courses. Especially students with an ambitious goal, who teamed up early with colleagues and fellow students and reached out early to people involved in the challenge to solve, scored high on impact. People with lower ambition levels seemed to be less driven to team up and go out. Based on their ambition, several students without programming experience even became programmers above the average group level within half a year. Driven by ambition seems to align with successfully driven by data.

Concerning the type of education there are three options: online courses, onsite (classroom) courses or participation in study teams. Especially the mix of these three showed to be the most successful, based on the study mentioned above. Next to these three delivery types, there is also an important choice on content to make as there are, on average, two options to pick from. :



JADS Marienburg

1. First learn about the data science techniques and then try to find related challenges (technology based).
2. Take a data related challenge and find out what is the best technologies to solve the challenge and learn to master the technologies while solving the challenge (challenge based).



250+
professionals

This last option has turned out to be the best way, after educating 250+ professionals in the one-year program of Professional Education at the Jheronimus Academy of Data Science. Where the focus is not on not just learning about Data Science, but on translating Data Science into Data Entrepreneurship which is the ability to create impact and value by discovering and exploiting new opportunities in business and society, based on new insights from data.

Technology driven education in Data Science

There is an enormous amount of technologies which can be used to perform Data Science. This includes programming in Python or R, but also database technologies, data cleaning, modeling, visualization and last but not least: mathematics and statistics. However, there is a risk: once you know and control a certain technology, you try to solve as many as possible challenges with that technology and you will slow down in learning to master other technologies. Which leads to the risk of applying a great technology for the wrong situation.

Challenge based education in Data Science

One way to solve the risk of focusing too much on one technology is by starting with the right framing of the challenge and choosing the best technology in the second step. This includes asking several technology specialists about the advantages and disadvantages of certain technologies for a challenge and testing several techniques, before choosing.

What are the biggest benefits of challenge based education?

First benefit: very motivated participants. As soon as well motivated professionals, who have chosen to spend one year long a full day and



Figure 1. The concept of Challenge Based Education

two evenings a week on data science, can work on their personal and business challenges, the heat is on. Both because there are so many tools available online and there can be such an open collaboration spirit between data science people. The statement “the best data scientist is a team” represents in a very clear way the DNA of good data scientists.

Second benefit: great networking in the course. At the start, when the fundamentals of data engineering, data analytics and data entrepreneurship are covered, there isn’t that much knowledge yet to realize the challenges. But thanks to a great mix of participants, there are always people who can help you to make a start.

Third benefit: great networking with colleagues outside of the course. When you can work on your own challenge, you have a great reason to involve colleagues and maybe even customers and suppliers. In this way, you get the opportunity to share the new acquired knowledge with several colleagues and customers. This contributes to another lesson learned: “teaching (to colleagues and customers) is the best way to learn”.

Fourth benefit: highest score on learning impact. When you work on something which is of real interest to you, you make a few extra steps on the learning ladder. Instead of just listening, and maybe being able to repeat, you learn to apply your knowledge and you even are motivated to add extra materials from your personal network or the online data science world.

What is the risk of challenge based education?

Of course you can only start working on your challenge when you master the basics. Your challenge should not lead to a situation in which you pay little attention to the basics and get of track. However, to prevent this, we have added practitioners to the course and the concept of buddy teams. This is done to avoid another possible disadvantage: risk of addiction. Based on the benefits above, there is a risk that great ambitions lead to some kind of addiction. E.g.: when you get support to solve your own challenge, you may feel obliged to also help the people who helped you. It did even happen that program participants instead of spending one full day and two evenings a week on the program, they started spending more evenings and also days in the weekend. This risk is partly avoided by introducing online hackatons during the course, for both the total group and for buddy teams. Thanks to this type of online teamwork, including online break-out rooms, you can solve more challenges in less time.

Results?

By adding challenge based education, including the buddy teams, practitioners and online hackatons, the participants have been able to focus more on creating impact with data science, next to learning the fundamentals. Almost half of the participants did save more for their sponsor than the cost of the course already before the end of the course. Some even realized a return on investment of more than 1000% (10 times the

costs of the program). Next to that, the participants satisfaction has gone up to more than 8 on a scale of 10. And above all: we are proud of all the participants who have become JADS ambassadors, JADS data entrepreneurs and JADS practitioners. As they have learned: teaching is the best way to learn, especially when it is combined with great teamwork and by working on real business or society challenges.

Want to learn more?

One of the founders of JADS, the Eindhoven University of Technology, has a well known centre of expertise for challenge based education, called the TU/e Innovation Space. To learn more, visit www.tue.nl/en/tue-campus/tue-innovation-space/

Practical and inspirational

"Last year I started with the Data Science for Professionals program. The program delivered on all points what I hoped for: knowledge, contacts, practical skills and a lot of inspiration. I can wholeheartedly recommend this program to any entrepreneur or professional who has ambitions around data science or artificial intelligence."t.

Anton de Nijs, Founding Partner, BrainStax

Step 2: Master the main data science technologies

Data Science is a science in which a huge number of technologies come together. It is also a science which is related to many other sciences and technologies. The picture below shows just a few of the data science core technologies and the relationship to just one other science: process science. Developments in Machine Learning, as part of Artificial Intelligence, may have had the most impact on the growth of Data Science in total.

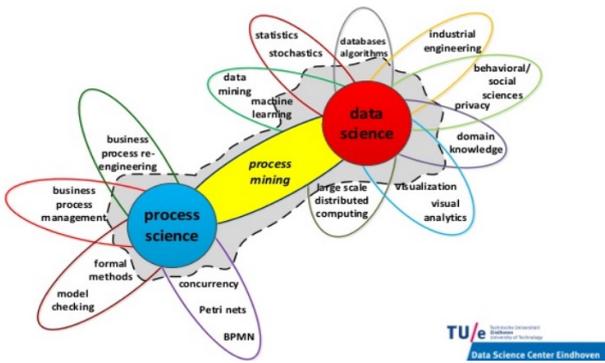


Figure 2. Overview of Data Science, Process Science and related science topics

For a more detailed overview of the technologies involved and the grouping in sub sections, see the figure below.

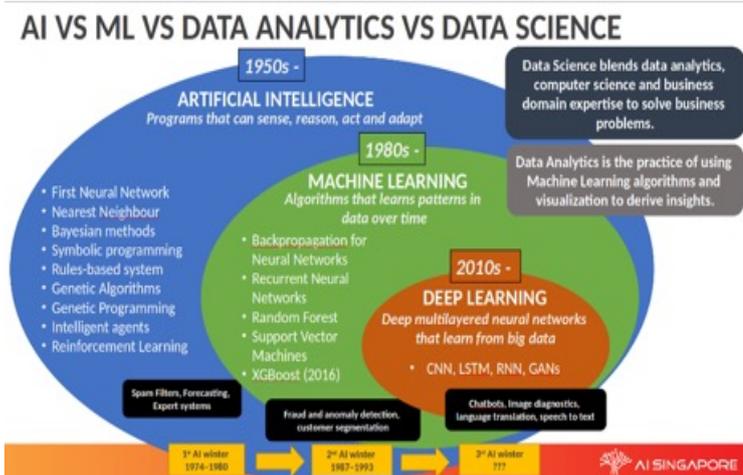
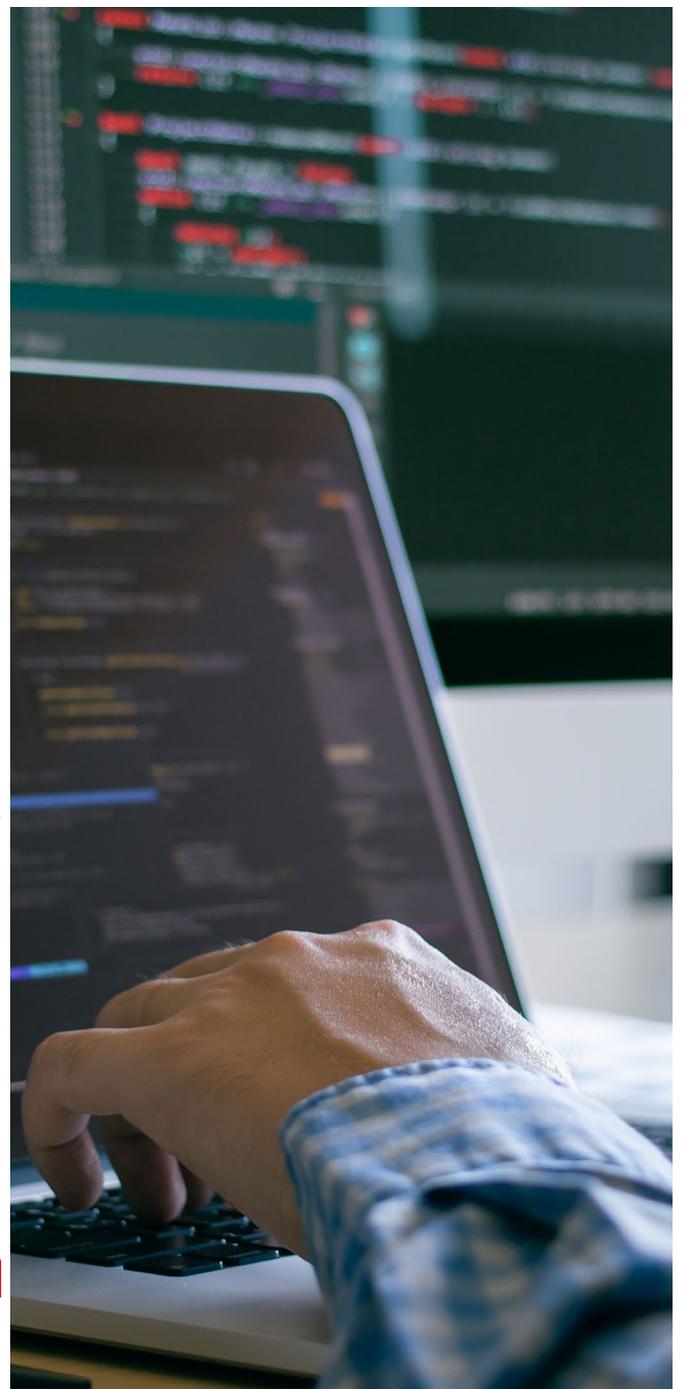


Figure 3. Example of an overview of Data Science which includes Artificial Intelligence, Machine Learning and Deep Learning

What are the main technologies to master? When we have to pick one main technology with impact, we pick Machine Learning. Machine Learning involves regression, classification and clustering technologies, which are the most used mathematical concepts.

Main sources to learn more about the main technologies are:

- [Datacamp.com](https://www.datacamp.com) and realpython.com for a nice range of online exercises
 - towardsdatascience.com for great articles
- See also towardsdatascience.com articles



Step 3: Choose the right project approach

Before starting a data science project it is important to choose the right methodology for your project. There are different methods you can choose from. At JADS we have opted for the CRISP-DM method. What are the other options, why do we prefer CRISP-DM and where can you learn more about this method?

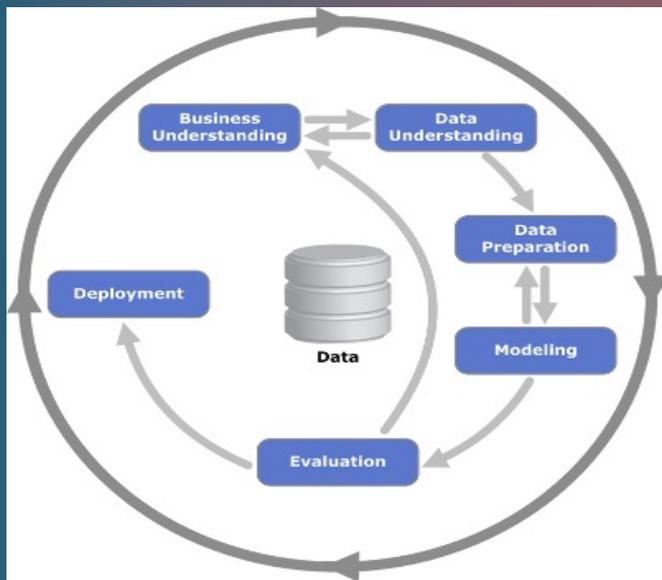


Figure 4. A CRISP-DM process diagram showing the relationship between the different phases of CRISP-DM. [Source](#):

What are the best known Data Science methodology options?

Next to CRISP-DM, which stands for cross industry standard process for data mining, there are several other methodology options, such as Scrum, Kanban and Microsoft TDSP. Below you can see the outcome of a recent poll on the popularity of data science methodologies. CRISP-DM is a clear winner.

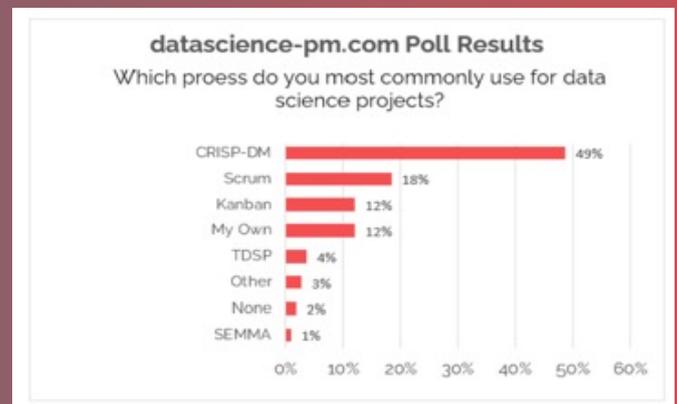


Figure 5: most common used data science methodology

[Source](#) To learn about more details and comparisons, visit this page: www.datascience-pm.com/

Why opt for CRISP-DM?

Next to the fact that CRISP-DM is the most common used methodology, many have opted for CRISP-DM because of the structured approach and the fact that it covers in a clear way the most relevant topics to pick from.

Maybe there could have been more attention to the decision making topics in Data Science projects, both at the start (“what decision making did trigger the topic for a data science project”) and at the end (“how to embed the insights found in the decision making”). But this can easily be added to the CRISP-DM methodology. See also the J and U model for data science projects for understanding the importance of analyzing the decision making process before implementing new data driven insights.

How to learn more about CRISP-DM?

To learn more about CRISP-DM, visit this [page](#). Another great source to learn more is this [guide](#).

Step 4: Secure your impact with focus on data driven decision making

For some data science professionals, data science starts with data engineering (cleaning, aggregating data), followed by data analytics (to find insights) and then testing and implementing insights found. We summarize this approach below as the J model approach for data science. However, since in this approach the insights are coming bottom up from already available data, such insights might be based on incomplete datasets and not resonate with the actual decision making challenges, processes and goals of decision makers.

We argue that a good data-driven decision making process should not start with data but with an assessment of the current insights and

issues, similar to the process model of Mintzberg (1976) and its related decision making. For this reason, we propose the U model as a more complete methodology for Data Science. Both the J model and U model methodology for data science are visualized in the figure below.

The first and final step in the Data Driven Decision Making (DDDM) process based on the U model for data science are being considered as important steps towards the success of data entrepreneurship. The U model shows in a clear way the importance of making the connection between DDDM and entrepreneurship.

In the first step of the U model the problem or opportunity is being defined, including a description of the current decision-making data and processes. We would like to describe this as the entrepreneurial or management decision making effort. In the second step it should be defined if and how data science can solve the issue or capture the opportunity. We would like to describe this as a part of the data entrepreneurship effort.

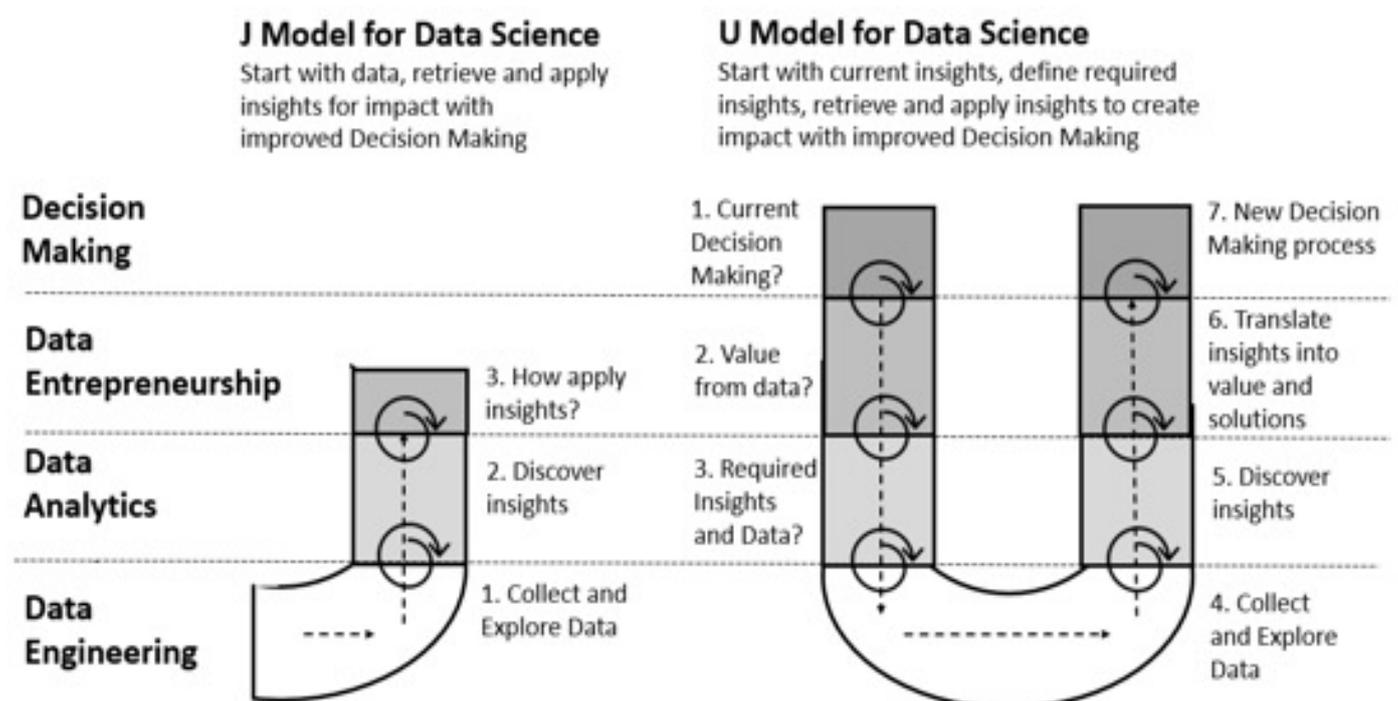


Figure 6: The J and U Model for Data Science - Source: Ronald Buijsse and Martijn Willemsen, Jheronimus Academy of Data Science (JADS), 2020

In the third step, the required insights and data should be defined, to guarantee that the engineering and analysis is done on the right dataset. In the fourth step, data engineering can start to collect or create the data required to discover the required insights.

What data is analyzed is driven by the opportunity or problem and the insights it requires, not by the data which has been collected for other reasons in the past. The fifth step in the U model aligns with the second step in the J model and contains data analytics activities to discover relevant insights. Step six is all about data-translation: when, where and how can discovered insights be translated into solutions and opportunities, among those as defined in step 2. Step 7 is about embedding the discovered insights into the decision making environment which was defined at the start, or removing the insights that were used, but proved not to be valid. This activity we would again describe as part of the data entrepreneurship effort.

The U model approach for data science thus extends the more traditional J model for data science by starting with the real problem, diving into the required insights and data, and after that following the J model approach to arrive at new, well implemented data-driven insights for appropriate DDDM.

Summary: How to develop yourself with Data Science

Data Science is all about discovering new insights in data, to make better decisions. Much data is already there: about your company, your processes, your customers, your environment. By applying the 4 steps in this whitepaper, data adds another view to your personal views. The data will challenge you and make you sharper. Educating yourself, staying on top of the latest technologies, choosing the right project approach and creating impact can make you into an expert in Data Science.

For more info on the Data Science for Professionals Program at JADS, visit our [website](#)

INFORM certificate (CAP)

“The Professional Education program of JADS turned out to be an excellent basis for obtaining the Certified Analytics Professional (CAP) certificate.”

Chris Niekerk- Lead data scientist Digital Advertising, DPG Media. k