**Data Science – Hiring thoughts**

Last week I had the opportunity to give an improvised talk in front of a young data scientist audience. I realized that talking on the spot helped me to structure my view of what skills should have a new joiner, and just thought on publishing the quintessence through a quick note.

Data science is an emerging topic, considered as the “sexiest job of the 21st century” by the Harvard Business Review. As a young field, qualifications are not clear now for the companies looking to hire / start a data science team if they lack analytical expertise. Majority of practitioners have learned their skills aside, rather than on a classical academic path: from a Human Resources perspective it can be hard to distinguish.

There are two dimensions to consider: the skills and the tasks location within the information pipeline.

We can split the overall skills of Data Scientist 3 main area, listed below:

- Machine Learning, that includes mathematics, statistics and algorithms knowledge

- Coding and software engineering

- Industry knowledge, specific for each available position

Figure 1: The 3 core skills

Maths / Machine Learning

Industry Expertise

Coding / Software Engineering

Majority of candidates will have strong skills in 2 of the 3 areas, with the specific profiles:

* Software / Maths: profile typically found in technologies companies, they started their career with a Bachelor or a Master in Computer Science
* Maths / Domain: they are traditionally researchers coming from the industry, most likely holding a Master or a PhD
* Domain / Software: this is less common, but candidates were often involved in data pipelines processes in the companies they worked for. They tend to have a technical background coupled with a business qualification, type MBA

There is then a second dimension to consider building the team: the place of the data scientist within the data pipeline.



Figure 2: The information pipeline

The role and the tasks associated will depend heavily on what is the objective of the future hire, and will condition the languages / tools to be used. Understanding its data flow is critical to then reinforce it: is the team more looking for a database guy or for a machine learning expert ?

This segmentation gives insight on how a company can effectively set up a data science team by pairing smartly candidates with complementary profiles, without the need of a large budget.

Each profile needs to be tested through customized questions regarding its strength, and the ability to work with the other profiles within a team. And the rest of the interview can be tailored to fit the specific needs of the company.