

Automation and its impact on Predictive Analytics – Creating the Analytical File (Part 2)

Written by Richard Boire at Boire Analytics

In my last article, I discussed the increasing impact of automation on business and the displacement of jobs. With artificial intelligence looming as the ultimate disruptor, the overall theme of job displacement has shifted more towards knowledge-intensive jobs which would of course include data scientists. The article ended on the note that this article would look at the future of the data scientist in a world increasingly more influenced by automation and artificial intelligence. In this second article, we want to focus its impact on the analytical file which represents the information inputs into any predictive analytics solution.

The year is 2019 and you are a newly hired junior data scientist where your primary role is to build predictive analytics solutions. What would you be expected to do? In order to better understand and appreciate the current role of the junior data scientist, it is important to understand what the role might have looked like in 2019. In 2019, the specific programming demand would have been for individuals with knowledge in R, Python, SAS, or a number of other more traditionally intensive computer-based languages such as Java or C++. High levels of mathematical and statistically-based knowledge would also be expected to be one of the core skills requirements for a junior data scientist. No question that the key skills requirements would have more of a technical bent rather than the softer skills which might translate to business knowledge and how these technical solutions would be applied in a business setting. The thinking in 2019 would be that the tech skills are the immediate need. Meanwhile the organization's training and internal development programs would build those softer skills of domain knowledge and how to practically apply these solutions within the given business. The development of this hybrid would be ever-evolving with tech skills as the initial foundation complemented by increasing domain knowledge. The more successful hybrids would comprise those data scientists who would ultimately end up in executive-level positions.

Now let's forward to 2029 and what might be the requirements of the junior data scientist. In an age of artificial intelligence and increased automation, the need for coding and programming will be minimized. But what does that mean to the junior data scientist of 2029? Information and data will still be analyzed and the need for an analytical file will remain as one of the core steps within the data science/data mining process and certainly in the development of any predictive analytics solutions. Yet, it is the tools which will improve in order to enable the data scientist to create the analytical file. We are already observing evidence of this through a number of vendors that offer GUI interfaces where the user clicks on icons that represent a certain data function. At the end of this process, the user ends up with a map of all the different processes and tasks which were required to create the analytical file. The analytical file can then be used to develop models or to produce reports and tables. There is no need for programming as the tasks and functions in creating both the analytical file and the required reports/tables are represented by drop-down icons.

This ability to facilitate the creation of the analytical file process is and will continue to be a key deliverable amongst software vendors moving forward. As a result, the data scientist needs to spend less time in the creation of programming code and more time focussing on aligning the right data to solve the business problem. Gone are the days when the data scientist would simply extract all the data. In a world of Big Data, access to data is no longer the challenge. Instead, the challenge for the data scientist is to be focussed on the data that will be relevant and meaningful in solving a given business problem. For example, if I am building a claim risk model, how relevant is the social media commentary related to insurance policies. The relevance in building a predictive model would only be significant if we can match a high portion of these policies back to their social media commentary. Yet, in another example in looking at insurance fraud, the use of social media might be used to detect patterns of communication where evidence of fraud seems to be most relevant. In both these cases, a stronger level of domain knowledge would help to guide the analyst in what data would be most meaningful.

The human technical skills of the past in creating the analytical file can now be augmented by software with the data scientist now focussing on the business problem, the approach to solving the problem and of course using the right data. The emphasis for data scientists will be on thinking rather than coding. As a result, more business challenges can now be tackled which in the past may not have been addressed due to data limitations or programming resource constraints. But data science in 2029 will still require that the individual have deep knowledge on data and how it can be used for data science projects. For example, the wide array of procedures and tasks that are used when manipulating data are core areas of knowledge to the data scientist. Keep in mind, data manipulation is typically well over 80% of the data scientist's time within a given data science project. This will not change in 2029 but tools will allow the individual to do this quicker thereby allowing more data science projects to be undertaken. Although programming will be less of a need, familiarity with the use of analytics software as well as data will be a core requirement. Better tools allow the data scientist to focus more on thinking through the problem rather than on programming. In other words, what kind of analytical file needs to be created to solve the given business problem.

At the academic level, colleges and universities will need to emphasize more of those softer skills in training students how to think through a given business exercise. Emphasis on courses with case studies will comprise a large portion of course outlines in any data science discipline. Yet, there will always be a need for those more hard-core technical programmers who can write code and the algorithms to solve problems which might not be solvable when using a more GUI-based analytics software. Some academic institutions may in effect design two tracts where one tract is geared more towards the creation of these data science hybrids as discussed above and another tract which emphasizes the more technical aspects such as programming and mathematics. In both tracts, though, the core ability of creating the analytical file is a key discipline outcome.

With academic institutions continuing to evolve their programs alongside the internal training and mentorship programs provided by many organizations, data science as a profession in 2029 will indeed be bright. Many new developments, techniques and approaches will emerge which is a natural outcome of our discipline. The use of the hard core technical data scientist in discovering these new developments complemented by the hybrid data scientist ensures that we are always looking at new solutions but with a view on how they provide incremental value over the status quo. In the next article, I will look at how data scientists in 2029 both develop and use predictive analytics solutions once the analytical file has been created.