

Automation and its impact on Predictive Analytics – The increasing importance of the hybrid (Part 3)

Written by Richard Boire at Boire Analytics

In my last article, I discussed the increasing impact of automation and its actual impact in creating the analytical file. As any data scientist knows, this component or stage of the data science process can typically represent well over 80% of actual project time with 90% not too being too atypical.

With artificial intelligence (AI) 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 more automated and artificial intelligence dominated world.

In order to better understand and appreciate what the future role of the data scientist might be, it is important to understand what the current role looks like in 2019. In today's current environment, specific skills include programming skills in R, Python, SAS, or a number of other more traditionally intensive computer-based languages such as Java or C++ which are the key object-oriented programming skills within the internet . High levels of mathematical and statistically-based knowledge also represent key core components of the data scientist's acumen. No question that these skills and competencies have much more of a technical emphasis rather than the softer skills which might translate to business knowledge and more importantly the application of these technical solutions in solving specific business problems. The thinking today is that the data science technical skills are the immediate need which would be complemented with domain knowledge of the business practice leaders in solving a specific business problem. But the operative word here in the subsequent sentence is "hope" in that we hope this complementation of both technical and business skills to be seamless. Yet, this is not the current business reality where in effect there is a gap between these skillsets resulting in a need for hybrids. In today's business environment, huge demand exists for these hybrids that can "bridge the gap in aligning technology and data to solve business problems. Organizations attempt to address this gap through training and "experience" as data scientists work closely alongside the business stakeholders. The development of this data science hybrid would be ever-evolving with technical skills as the initial foundation complemented by increasing business domain knowledge. Those data scientists who gained more knowledge and expertise in the business domain knowledge area would have much more career potential in moving into the C-suite. Yet, although there is still this void in the current marketplace for the more data science technical skills, the need for the hybrid continues to increase as data analytics is now a core asset for all organizations.

Now let's forward to 2029 to examine the requirements of the junior data scientist. In an age of artificial intelligence (AI) and increased automation, the need for coding and programming will be significantly minimized. Technical skills are less of a requirement. Apps and software are readily available to be used by the data scientist thereby mitigating the need for deep technical skills. The two key core requirements of a 2029 data scientist will be process and interpretation.

Process represents the exercise of “working the data” in order to create the appropriate analytical file which would ultimately lead to the development and implementation of a solution. Meanwhile, interpretation represents the ability to understand output and results that could arise from both complex mathematical algorithms as well as simple reporting exercises. Interpretation, however, does not require that the data scientist need to know the arcane mathematical computations used to develop a certain mathematical algorithm. Tools are now available which require minimal knowledge of programming and coding in “working” the data. At the same time, open source applications as well as commercial applications offer a variety of advanced mathematical and statistical routines which can be executed by the data scientist. In this type of environment, less knowledge is required on the arcane mathematics that is used to actually create a mathematical solution. Instead, the emphasis is on the more “analytical” thinking and the interpretation of results as well as the ability to effectively process the data into an analytical file.

Having said that, there will always be the need for a group of hard-core data science programmers which would represent those individuals who push the frontiers of data science in developing new algorithms and solutions. We have witnessed examples of this “push” through the latest developments in AI in the last 5 years that have led to significantly better solutions in such areas as image and text recognition. Supplier organizations that offer advanced analytics solutions including AI in the form of software, apps, etc. would certainly have these typical “hard-core” data science programming type skills. Meanwhile, the end-client companies within the many industry verticals would have the more typical “hybrid” data scientists. These hybrids would need to be comfortable in applying more advanced type AI algorithms but with a clear understanding of its impact in resolving a given business problem. But how does this shift occur in evolving data science to be more business-oriented.

Clearly, it would commence within the academic community. Certainly, the more traditional data science programs would continue to offer the more technical components of data science under the auspices of the faculty of computer science or mathematics. But the much larger need for the hybrid data science programs would be under the faculty of business where the goal is to create business practitioners of data science. Aligning the right data to solve the right business problem would become the core foundation of any data science hybrid program. Organizations would then provide the practical business experience to further complement this academic learning.

But within this brave new world of 2029, the clear emphasis would be on the data science hybrid. More emphasis would be on analytical skills that go beyond the limits of technology including artificial intelligence. Non-linear type thinking and “out of the box” type thinking is now required for the myriad of business problems and challenges that can now be addressed which were otherwise unsolvable due to technology limitations. In this brave new world, technology and artificial intelligence is simply an enabler while data scientists are now the problem solvers.