

Summarizing “A Few Useful Things to Know about Machine Learning” by Pedro Domingos.

Pedro Domingos, also commonly known as the author of *The Master Algorithm*, wrote an article on a few key lessons to keep in mind when practicing techniques in machine learning. It's worth noting that all of these points are discussed in greater detail under the class *Machine Learning* (MSAI 349) at Northwestern University, taught by Doug Downey. Domingos goes into explicit detail on what methods in machine learning can further reinforce procuring more accurate results, and what erroneous, habitual practices have led data scientists to misinterpret or fail to obtain desired outputs. Some of these suggested practices include but are not limited to: paying close attention to the components of learning, the need for inductive bias, combatting overfitting, the curse of dimensionality, multi-model training, Occam's razor (the simple algorithm may be the smarter algorithm), and the golden rule of research design (*correlation does not imply causation*).

Unlike the more traditional research papers we've read until now, this informative piece represents guidelines to obey. The biggest strength of this paper is the content itself, and how these are some of the most important lessons that ML practitioners have honed with time and cannot be found easily in any textbook. A second, very important strength is that while this paper details how and what to do when studying ML, it also discusses what pitfalls and obstacles scientists have come across before and failed spectacularly in before. This may in fact be more helpful than the former because obeying these will, at worst, keep you away from the spiral of making the same mistakes others have gone through before. Examples where Domingos teaches us where errors lie appear throughout the paper. One such would be under section 5, saying “A common misconception about overfitting is that it is caused by noise...”, and again in section 10 when discussing the sheer number of learners: “In the early days of ML, everyone had their favorite learner, together with some *a priori* reasons to believe in its superiority.” This warns readers even before we discuss the topic of multi-model learning that sticking to one had its repercussions. Lastly, this paper is coherent enough to be a great read even for the young programmer. Some discussions are even written in layman's terms or given representations and analogies to keep all audiences on the same page (pun not intended).

I wasn't particularly a fan of the author's use of jargon in this paper. Before we leave the abstract, the author already makes use of “black art” when talking about applications in ML, which I still haven't been able to wrap my head around. It really doesn't help to google search “black art” either, or else you get something [like this](#). Additionally, as the paper has only the length to glance over each topic, it doesn't do so well to provide great depth in the form of more illustrations or example problems. Most of his discussion is conceptual. Finally, it seems like he cut the paper short at conclusion and chose against going over more points such as referencing other studies/articles because he wanted to promote his book as the only tool to continue forward with this article. His real conclusion is also only 2 lines otherwise.

So, while there were some flaws in this article in the author's take on the paper, the fact that the paper is very coherent for all to read, that it builds upon each topic effectively, and even

makes use of illustrations to represent techniques of ML makes it a great source for all practitioners to read before they immerse themselves into this world. As for future directions, I'm aware Domingos already has a book published and hinted to read it at the end of this article, but otherwise he's done great at providing a non-technical introduction to ML programming. I've also heard that he teaches at U Washington, but I'd suggest he travel to other universities to lecture and impart his knowledge to others. I know I've seen his book in lecture slides by other professors, so I can only assume his work has significant influence with some researchers here at Northwestern already.