Reading Response March 30

Kritika Chugh
SUID: 882046659

kchugh@syr.edu

The paper pointed out an important yet unexplored area in AI and ML models – AI Explainability. As pointed out in this article, we live in a world of algorithms defining us and yet scientists working in the field have talked mostly around ML interpreting over the explain ability. The paper spoke of XAI in a very comprehensive way talking about the areas it affects and the way it can be achieved. What I like mostly is that they also talked about why this being an important (unexplored area) is still not popularly practiced. As we know that there exist multiple accurate models that can be achieved using same ML algorithm but taking a different internal pathway every time. This gives us many accurate models and thus the explanation varies. Models having multiple layers are more intricate and difficult to both interpret and explain. And may be that's why XAI is not used by everyone. But I also feel, that explaining their models even after they are doing the job, they are supposed to, would deter these people exploring the explainability of the model. What I feel is that there exist less investment-both monetary and motivational that causes this problem. Furthermore, AI is ubiquitous, participate in many critical life changing decisions and therefore I personally welcome XAI into AI landscape. The goals are to make AI and ML more explainable to the humans, transparent, demystifying the black box and in compliance with the legislation. So, in order for humans, organizations and even government to trust the AI systems Explainable AI is has to be incorporated into the models. **Transparency and Expalinability** are going to be the new buzz words I guess in the world that sells these AI systems to the end users and what not to like about it?