Thursday: Ethical Implications of Supervised Learning

Agenda

5 min: Overview

50 min: Suggested Resources

1 hr 5 min: Exercise

Specific Learning Outcomes

- I can identify ethical issues associated with applications of supervised learning in a variety of professional settings.
- I can assess the actions of individuals, corporations, governments and other organizations as ethical or unethical.

Overall Learning Outcome

- I can identify the regulations, standards, and ethics surrounding predictive modeling.

Overview

During the Moringa Data Science Prep we learned how sometimes as Data Scientists we can get lost in the techniques and methods of our trade.

In this session, we're going to learn that sometimes we might forget to ask important questions even during our modeling and deployment. We might ask a question such as "Should we include features that could be discriminatory?".

A good example of where we might fall prey to wrongful assumptions can be in cases say where features such as address code and tribe/race have historically had a disproportionate effect on loan access and interest rates for minorities. To learn more about where mistakes have been made during the course of a data scientist's work, let's now go over the following activity.

Activity

For this activity, we will be required to go through the following readings then use the following checklist **[Link]** **(https://docs.google.com/document/d/18Hkb8ojA7ZNiUEgaBrkyNzWW9c2EXZVR0eQKuhuq4BE/edit?usp=sharing)** to determine where each case study falls in the given checklist.

- Amazon Doesn't Consider the Race of Its Customers. Should It? [Link] (https://www.bloomberg.com/graphics/2016-amazon-same-day/)
- A Child Abuse Prediction Model Fails Poor Families. [Link] (https://www.wired.com/story/excerpt-from-automating-inequality/)
- Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. [Link] (https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#7e46da54713d)
- Research shows gender bias in Google's voice recognition. [Link] (https://www.dailydot.com/debug/google-voice-recognition-gender-bias/)
- Facebook Tweaks Newsfeed to Favor Content from Friends, Family. [Link] (https://www.wired.com/story/facebook-tweaks-newsfeed-to-favor-content-from-friends-family/)
- YouTube's Creepy Kid Problem Was Worse Than We Thought. [Link] (https://gizmodo.com/youtubes-creepy-kid-problem-was-worse-than-we-thought-1820763240)
- Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. [Link] (http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf)
- Why Google Flu Is A Failure. [Link] (https://www.forbes.com/sites/stevensalzberg/2014/03/23/why-google-flu-is-a-failure/#3dec576e5535)
- What happens when an algorithm cuts your healthcare. [Link] (https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy)
- Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech. [Link] (https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai)
- Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased? [Link] (https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/)
- Microsoft scrambles to limit PR damage over abusive AI bot Tay. [Link] (https://www.theguardian.com/technology/2016/mar/24/microsoft-scrambles-limit-pr-damage-over-abusive-ai-bot-tay)

"War is 90% information." – Napoleon Bonaparte, French military and political leader