BBM467 - Blog Post Project

Crocodile

21827555, Meltem Kaya
2210356157, Safiye Tüfekçi

The Hidden Biases in Data Science

If you are into data science, you have probably heard the phrase: "All datasets are biased, some are useful." When working with a dataset, firstly situations that may create bias in the dataset are considered and precautions are taken against them. But are all biases so predictable? The answer is of course no. These unpredictable biases are called hidden biases, perhaps the most dangerous of all biases. Sometimes the effect on the results is so little as to ignore, and sometimes it affects the results greatly.

Figure 1. Bias, bias everywhere

Let's start with a simple example for better understanding before diving into data science…

Remember the dress that many of us have seen on social media and that completely divided the world into two groups. Some people insisted that it was blue-black, others gold-white. Well… What colors do you see? *Here* is the link for those who cannot remember the image.

This optical illusion is called color constancy. This image just happens to be perfectly ambiguous to our brain, meaning it all comes down to individual perception and even how and where you view the image. Since we do not know in which ambient illumination this

image was taken, our brain completes this missing information on its own and interprets the image depending on the ambient illumination we are in while completing the image.

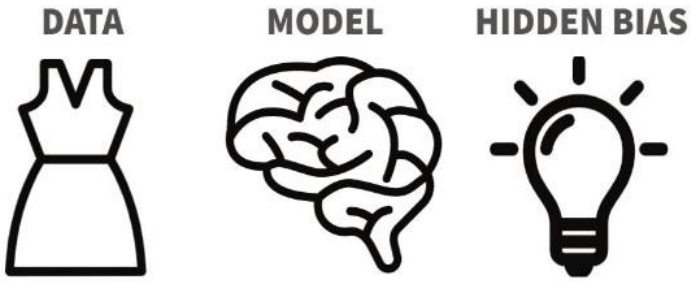

Figure 2. Inferences

To explain this sample in the context of data science, the dress image is data; our brain and its interpretation is the model, and the color we see is the output. The hidden bias here is the illumination.

An Interesting Example from Amazon…

You remember the "sexist" AI that Amazon trained in 2014 to streamline its hiring process. Amazon trained an AI model that scores job applicants with data from past years of employees. After one year, it turned out that artificial intelligence did not like women. It lowered the scores of women applying for the job and filtered out candidates who went to girls' high schools. This is because, the fact that most of the profiles used while training the model were male profiles. Although the engineers did not add such a gender parameter, the AI learned from the data presented to it that the gender of the employees must be male. In fact, after the engineers eliminated this situation, they thought that artificial intelligence could find new ways to discriminate against candidates, so they canceled this application.

Figure 3. Psychopathic AI

As it can be understood from this situation, an unpredictable bias in the data is hidden and sometimes affects the research results indirectly, sometimes directly and concretely, as we have seen in this example.

In Conclusion...

Big Data is this big, messy thing produced by a human, often one of the most complex systems we know. The typical approach to dealing with this complexity is to abstract everything into a black box or simply let the machine learning algorithms handle the complexity the best they can. But, we cannot be a hundred percent confident in the results we obtain by focusing only on the numbers in the data. The main thing here is to be able to

read the facts behind the numbers. The danger here is the fact that our model has already learned biases that we were not aware of.

References

[1] Crawford, K. (2013). *The hidden biases in big data*. https://hbr.org/2013/04/the-hidden-biases-in-big-data

[2] BBC. (2018). *Amazon scrapped 'sexist AI' tool*. https://www.bbc.com/news/technology-45809919

[3] AsapSCIENCE. (2015). *What colour Is this dress? (solved with science). https*://www.youtube.com/watch?v=AskAQwOBvhc

[4] Wikipedia. *The dress*. https://en.wikipedia.org/wiki/The_dress

[5] Data Science Conference, (2018). *Hidden biases in data: moving beyond black box models - Marko Jevremovic*. https://www.youtube.com/watch?v=orT8617DJJo&t=1016s