ADDRESSING BIAS IN MACHINE LEARNING APPLICATIONS

Many experts believed artificial intelligence (AI) and machine learning would usher in an era of efficient, transparent and sound decision-making. However, bias continues to be a problem because humans create the technology and incorporate their prejudices and unconscious beliefs into it.

Machine learning models learn what they are taught, based on the data provided. They use advanced algorithms to analyze vast amounts of data and recognize patterns to glean business insights. If the data is flawed, then the results will be skewed.

HOW BIAS CAN SKEW RESULTS

As machine learning quickly becomes a powerful tool for businesses in a range of industries, undetected bias can result in life-changing outcomes, such as a candidate being disqualified for a job based on the race or gender of similar past candidates.

Machine learning bias originates with human bias, preconceived ideas and prejudices. It can enter during the development of the machine learning application or in the data used to test the program. Biased results perpetuate stereotypes, trigger erroneous thinking, and misdirect projects with costly and damaging results.

Bias can creep into social media applications of machine learning. Algorithms learn from interactions among users and, as a result, they steer news, advertising and posts according to the user's likes, interests and contacts.

Google searches can steer users toward preferences instead of neutrally offering options. Facebook users may fail to see all their friends' posts on certain topics that the algorithm decides aren't of interest to them.

If not caught and corrected, this kind of interactive bias can steer perspectives in distorted ways. The options available to social media users can shrink to reflect a narrower perspective. Users may suddenly find fake news stories presented as just as valid as legitimate news sources.

Another problem with machine learning bias is that it creates synthetic data points to bridge data gaps for predictive modeling. These synthetic data points can miss conditions that were previously unimaginable and, thus, lead to erroneous conclusions.

For example, at the beginning of the Great Recession, economic forecasting models were not built for an environment with zero or negative interest rates. But such an environment emerged and the models had to adjust to the new reality.

Machine learning models learn what they are taught, based on the data provided. They use advanced algorithms to analyze vast amounts of data and recognize patterns to glean business insights. If the data is flawed, then the results will be skewed.

PNC

Once the application gains experience and improves performance, further investment can be made to scale up the technology. This enables companies to check for bias and tweak and validate a machine learning model in a thoughtful and balanced manner.

DEALING WITH BIAS

Data scientists are finding ways to minimize the impact of bias in machine learning applications. Several best practices have emerged to guide companies in their efforts to implement and expand their use of machine learning:[1]

- Although it may seem uncomfortable to openly and honestly ask questions about preconceived notions, prejudices and assumptions, this step is critical for companies applying AI and machine learning. Because it can be awkward to ask difficult questions about unconscious biases, many organizations use consultants to challenge the assumptions being made.

- Large sets of data are used to "train" machine learning models. This is where the old adage "garbage in, garbage out" applies. When companies carefully consider potential biases in the training data set, the potential for skewed results can be significantly reduced.

- Testing an algorithm can uncover unanticipated biases and skewed conclusions. This enables companies to adjust the type of data used for input or the algorithm itself to achieve desired results and avoid unintended outcomes.

- Increasing the amount of training data can reduce the potential for bias as well, as more data often offsets potentially problematic results by providing greater diversity.

- For predictive modeling, companies can overcome a tendency of the algorithm to revert to past outcomes by using experimental techniques to model potential outcomes.

- Because the world is constantly changing, machine learning models must be maintained and updated periodically. New data sets should be tested and used to reflect the changing environment.

As machine learning moves into the mainstream of business analytics, many data scientists advocate using a staged approach to developing and expanding algorithms. Once the application gains experience and improves performance, further investment can be made to scale up the technology. This enables companies to check for bias and tweak and validate a machine learning model in a thoughtful and balanced manner.

To discuss these topics in more detail, please contact your PNC Relationship Manager.

1   "The Risk of Machine-Learning Bias (and How to Prevent It)," by Chris DeBrusk, MIT Sloan Management Review, March 26, 2018.
    Available at: https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/

PNC