STANDARDS FOR RESPONSIBLE DIGITAL FINANCIAL SERVICES: ALGORITHM BIAS STANDARDS SECTION

2 May 2022

CERISE + SPTF
Draft Standards: Algorithm Bias
(as of 2 May 2022)

Standards for Algorithm Bias

1. The provider defines specifically what "fair" algorithmic function means.
2. The provider shares the definition of fair with its board of directors.
3. The board of directors hold management accountable for fair algorithmic function.
4. If outsourcing algorithm development:
   - Inform your development partner of target customers and discuss a strategy to avoid algorithmic discrimination.
   - In the service agreement, do the following:
      i. Define parameters for algorithm
      ii. Define what specific tests the partner will run to check that that the algorithm is "fair" according to your definition
      iii. Require the partner to check annually the algorithm's fairness, according to the definition determined by the provider, and make corrections, as needed.
5. If developing the algorithm in-house, credit officers and management take part in the development of algorithm design.
6. If you have information technology (IT) specialists developing your algorithm, train them on your mission and vision and target customers so they understand the context in which the algorithm will be deployed.

- Before you launch the use of algorithm, test whether your *data* are biased.
- Before you launch the use of an algorithm, use synthetic or real data to test the following:
   i. Whether the algorithm is "fair," according to your definition of fairness
   ii. Whether the algorithm is treating equally men compared with women
   iii. Whether the algorithm is treating equally any other customer segments that are relevant to your social goals (e.g., rural vs. urban)

7. Before you launch the use of an algorithm, take the following steps to solicit feedback from stakeholders:
   i. Identify the stakeholders involved in the use of this algorithm
   ii. For stakeholders who are not clients, speak with representatives from each of the stakeholder groups to identify any concerns they have about the use of the algorithm
   iii. For stakeholders who are clients, interview representatives from each segment of customer that the FSP identifies as important (e.g., women/men)
   iv. Document what you've learned in a way that makes it clear which stakeholder group had which concerns.
   v. Qualify risks in terms of which would be high or low priority to mitigate, and then decide which you will address and which you will not
   vi. Take action to mitigate the risks you are going to address
8. Analyze your algorithm function for fairness on an ongoing basis, according the frequencies below:
   i. If the algorithm is learning continuously, check function at minimum monthly.

ii. If an algorithm function is fixed, check function at minimum annually.

9. If you find that bias exists, determine if it is coherent with your social goals and strategy.
10. Prepare reports, at minimum quarterly, on algorithm function. Analyze at minimum this:
    - Who is being approved, by customer segment, and compare who is actually being served with the market that you want to serve
    - Whether the algorithm is accurate (e.g., check whether the algorithm's decisions on loan sizes for target customers are the same that traditional repayment capacity analysis would make)
11. Share reports on algorithm function with senior management, credit department, the risk management team, and the board of directors; discuss results and identify any corrective action needed.
12. Use information from customer complaints to inform your review of algorithm function.
13. In cases of a systemic shock (e.g., a pandemic), discontinue the algorithm and review it.
14. At least some members of the team that define algorithmic "fairness," and determine what analyses to conduct to test fairness, represent the population whose data are being scored by the algorithm.
15. Do not use algorithms if you do not have the capacity to analyze whether they are fair.

Guidance on mitigating algorithm bias-concepts, examples of real practice, and questions to address

Concepts

- AI and machine learning can be a tremendous force for inclusion, particularly for clients who are thin-file customers.
- We could also talk about "algorithm fairness." In other sectors, there has already been a lot of work on algorithm fairness that we can draw on.
    - One possible definition of fairness, shared in a WWB paper: "Statistical parity: Subjects in protected and unprotected groups have an equal probability to be in the positive predicted class
- No algorithm will be 100% fair. Also the definitions about what "fair" means differ from one FSP to the next.
- Accuracy does not equal fairness, though there is a lot of overlap. For example, people of color in the United States have a lower probability of having access to the internet, and an algorithm may accurately make predictions based on access-to-internet, but is that fair?
- It is not necessarily desirable to eliminate bias from an algorithm, as you may want to bias the algorithm to prioritize serving certain populations.
- FSPs are using data that are not directly related to credit history and credit demand or needs to make credit decisions. For example, how active a person is on a gambling app. Or where a person spends his/her time, as tracked by GPS data.
- Certain sources of data might be more subject to data bias. GPS is an example of this. Women tend to have more responsibilities at home whereas some algorithms use GPS to see how much time a person spends at his/her place of business. Another example relates to Smartphones. If more men than women have Smartphones, the majority of the data you extract from phones will relate to men and not women.
- WWB recommends, "Build a fairness implementation team. This multidisciplinary team should bring a group of legal, business, and machine learning experts together. Legal advisors define what the legal constraints are or could be, identifying what the minimum

threshold of compliance might be — and how to design for future regulation. Business experts think about what definitions of fairness fit well with their strategy."

- Although a third-party evaluation of an algorithm's bias could be useful, avoid requiring it, as the expense is not feasible for many FSPs.
- The probability that a third-party algorithm developer is going to show you its entire code is low. Do not require that third parties share this.
- If women have higher loan repayment rates than men, that likely means that the provider is not getting enough loans out to women.
- Data scientists and coders may live and work in a country that is far away and dissimilar culturally from the FSP that will use the algorithm that the data scientists write.
- Some examples of customer complaints that can inform your analysis of how well your algorithm is functioning: a) a customer complains she didn't get as big of a loan as her neighbor; b) a customer says it's been months and she hasn't been offered a larger loan size.

Examples of real practice

- FSPs that are actively managing algorithm bias whom we might approach for case studies: Maha (Myanmar), Jumo (Zambia), and Tala (Kenya)
- WWB is working with a provider that can see that among the prospective clients whose requests for credit are reject, the women have a higher credit score. WWB is putting together a scorecard that somebody in management could hand to their data scientist and say, "Give me how we score on all of these things."
- WWB is developing a scorecard with five common definitions of fairness. For example: a) different groups of people have the same repayment rate; b) study the likelihood of receiving a credit off; c) study the likelihood of being rejected for a loan
- "Women's World Banking recently created a Python-based toolkit to show how financial services providers can detect and mitigate gender biases in credit score models. The first step in the toolkit is a series of questions on portfolio size, sex ratios among customers, likelihood of women versus men applicants being extended credit, and a number of other factors. By asking these questions, the tool can model a particular institution's credit portfolio. Next, based on user input, the tool creates a synthetic dataset for the user and provides insight on both bias detection and mitigation. Visit the tool at github.com/WomensWorldBanking."
- ORCAA asks: For whom could this algorithmic system fail? This directs the conversation to thinking of all the stakeholders: credit officers whose job it is to make loans, there's the risk management team whose job it is to ensure high repayment, there's the mission team whose job is to care about the socioeconomic profile of clients, and outside of the FSP there are also stakeholders, most notably the clients, and also the vendor that sold the algorithm to the FSP.

Questions to address

- Is "Protected Classes" too much of a United-States law term to be the right phrase to learn internationally?
- What about groups who are not considered Protected Classes in markets outside of the United States, like certain ethnic groups, but tend to be subject to discrimination?
- What is the responsibility of the FSP to inform clients about how the algorithm functions and/or why the algorithm made the decisions related to the offer of products for clients?

- What if clients are uncomfortable with certain data the FSP collects? For example, using GSP clients to track their location. Clients in the CFI study were not comfortable with non-traditional data sources like airtime top ups, text messages, what apps people have on their phone, and battery usage being used to make decisions about whether to offer a loan.