Data Ethics and its importance in the Financial Sector

Ananya Jha

14/04/2022

Introduction - So what is data ethics and why is it becoming increasingly important?

The use of Big Data, Artificial Intelligence, and Machine Learning is constantly progressing. The applications are extensive: from day-to-day applications to complex analyses and the benefits are just as vast. The future looks bright as many AI platforms like Deepmind and Watson are looking to now be made available publicly. As their applications and scope expand, there are also many issues arising such as data privacy, unfairness (or bias), and lack of transparency in the system. This branch of ethics that discusses these moral issues associated with data is known as data ethics.

There is no given set of rules, agreed definition of data ethics or an oath (such as the Hippocratic oath for doctors) and the areas are mostly gray. But as statisticians, it is our moral obligation to ensure our models are fair are representative. Merely thinking about what is possible for us to do with data is not enough. We need to think about what should be done with data.

Ethical issues might be either data collection related (like data leaks for health data, the existence of indirect identifiers which violate privacy, intrusive surveys) or related to making decisions with data. There are various examples of unintentional bias in the system that arise from reproducing already existing biases. Like the use of AI in predicting if criminals will be violent or if they will resort to criminal behavior when released has seen major differences in the rate of false positives for people of color. AI based recruitment algorithms have seen biases against gender and the use of Standardized Testing Scores have also proven unfavorable for some groups. In this paper, we will be exploring the making decisions with data part with a specific focus on the financial industry.

Data Ethics in Finance

The use and applications of AI and ML has been expanding profusely in the finance industry and AI is considered the future of the banking industry. It is more powerful, makes processes faster, can process more data than traditional methods and include more variables and so consequently is more accurate. It is also very cost effective. In fact, a recent report by Business Insider suggests banks can save approx. 447$ billion by incorporating AI in their front, back and middle office channels(Digalaki, 2022). It is being used everywhere from front offices like virtual assistants to help clients, to detect fraud/illegal activity, for compliance monitoring tools, in trading and in predictive analysis for credit underwriting and fair lending.

It is important to be ethical in every industry when working with data but in finance specifically, the repercussion of unethical AI is massive. Denying critical opportunities to minority groups can significantly delay development of the economy, greatly harm certain groups, and lead to distrust in the specific Financial Institution. Consequently, if an FI is deemed to be unfair, the costs of declining public reputation and legal fees are far too great.

The importance of Explainability and Interpretability

Explainability and Interpretability are considered two important pillars of ethical AI. Understanding their definitions is the first step to being an ethical AI practitioner.

"*Interpretable* usually describes directly transparent or constrained ML model architectures, and *explanation* is often applied to a post-hoc process that occurs after model training to summarize main drivers of model decisions8. Both concepts are important for adverse action notice reporting, because the more interpretable and explainable an ML system, the more accurate and consistent the associated adverse action notices(Orrell & Houshmand, 2021)."

Intuitively, it should be critical to explain the decisions made by AI like why a person was denied a certain loan or mortgage or why their credit score is low or even high. Not only is it important for the clients to be aware of this, it is also important for the regulators especially if the AI is wrong or biased. Moreover, understanding the models is also important for further advancements and improvements to increase accuracy, increase efficiency and reduce biases.

Presently, most applications of AI in finance value accuracy and precision first and consider explainablity or interpretability an afterthought. For e.g. Credit score algorithms and credit risk models are considered black-box. There is a major trade-off. The more complicated the model, the tougher it is to explain to a non technical audience. The very reason that ML is being used is because the problem that we're trying to solve is complex so obviously it is going to be hard to explain.
However, many companies are trying to build models that reach the optimal level of accuracy along with interpretability and explainability. Models like Linear models(and its variations like GAMs), Neural Networks, Decision trees achieve this feat. Another way to tackle this issue is with Post Hoc explanations by creating summaries that either partially or fully explain model behavior(Orrell & Houshmand, 2021).

The Problem

Over a long period of time seen that financing system has been less favorable for people of color. Initially, AI was expected to remove bias that comes from human judgement but that has not worked well so far.

The primary point to understand is that Machine Learning is data driven. It makes accurate predictions when the future looks very similar to the past and there is lots of reliable data from the past. Good models also aim to provide high accuracy. By design, a Machine Learning model reduces predictive mean squared error which leads to models that have greater variance.

Essentially, AI finds a model that fits the majority of the population even if it doesn't work well for minority groups. In this case of predictive analysis, the past data distribution is not the same as the present and that causes more ethnic and racial discrimination.

To further understand the cause and how Financial Institutions are trying to fix it, we need to define a few terms first.

- **Protected characteristics:** A protected characteristic is basically a personal trait that are protected from discrimination and harassment by law and hence must not be used as a variable in an ML model. Examples include(but are not limited to) race, ethnicity, gender religion, nationality etc.
- **Proxy variables:** Attributes that are not a direct measure of certain variables but are highly correlated to them are known as proxy variables. In this context, attributes such as zip code can be proxies for ethnicity. Often, these are not that obvious. Things like majors or certain shopping history (like shampoos) are highly correlated to protected attributes like gender.
- **Disparate treatment:** "Disparate treatment occurs when a lender treats an applicant differently based on one of the prohibited characteristics in any aspect of a credit transaction, including the provision of credit and setting of credit terms (e.g., pricing). It is always illegal in lending and does not require any showing that the treatment was motivated by prejudice or a conscious intent to discriminate" (Orrell & Houshmand, 2021).

- **Disparate impact:** "Disparate impact occurs when a lender employs a neutral policy or practice equally to all credit applicants but the policy or practice disproportionately excludes or burdens certain persons on a prohibited basis. Disparate impact is not necessarily a violation of law and may be justified by a business necessity, such as cost or profitability, and by establishing there is no less discriminatory alternative to the policy, practice, or model" (Orrell & Houshmand, 2021).

Commonly employed solutions

Naturally, with trying to fix these biases, the trade-off between accuracy and fairness is of the utmost concern. The key is to find the maximum accuracy that you can achieve while being as fair as you can.

Financial Institutions have employed various ways to try and fix these biases without severely affecting the accuracy of their models.

Prior to building the model

Before building the model, the focus is on the data. As ML is primarily data driven, it is important to have better, more recent and representative training data. Ideally, the data should be complete and tested for integrity and accuracy. The data should represent an equitable world which might not exist in reality.

Due to long running historical bias, the data is largely unreliable. In many cases FIs use AI to first spot patterns of historic issues and alter this data to make the model more equitable. For example if it is observed that a certain minority group historically has to earn x% more to be approved for a loan, AI can be used to balance this by shifting the distribution.

During modeling

There is a process to regularize a model. If there is a noticeable discrepancy or unfairness in the model, we can score the model on a scale. Including an extra parameter if the model treats a specific protected class differently to fix that difference has often worked.

Another way to is to construct a separate model to predict if there is bias against minority groups/protected classes by the original model. These models are often called adversary models and can detect protected variables which helps find proxy variables. For example, a certain minority group might get different treatment due to a zip code being included in model. The adversary model will show that the particular minority group gets treated differently/gets lower limits even if protected variables were not used.

After constructing the model

Post development of the model, many FIs resort to different testing methods. A commonly used technique is matched pair testing. It is used to compare if treatment quality is equal for different groups, to identify disparate treatment and asses fairness. The idea is to match one non protected class to a protected class. Their profiles should be the same with the only difference that one is a minority group and the other isn't. If the model performs the same for both groups, it is fair. If not, further modeling and investigation would be required.

Another idea which hasn't yet been used in this context but might be very effective in finding the presence of proxies is a blind taste test. The way it would work is the model would be trained first by excluding the data from a specific protected community and then re trained including that missing data. If the accuracy is equally good in both cases, it would mean model is blind to the issue and there are no proxies.

Successfully removing biases from the system would break the self-perpetuating, self-fulfilling prophecy that existed in the human system without AI, and would keep it out of the AI system(Uzzi, 2020).

Why not completely remove AI and ML from this field?

Completely removing algorithms from this field is not the best idea. There is a reason why major banks like Capital One, Citi, HSBC, JPMorgan Chase are all employing this approach.

Sure, in situations like police work or criminal bias, since the cons far outweigh the pros, it might be better to depend more on humans but in cases like fair lending, credit score assessment etc, the advantages make ML, AI and Big Data worth it.

Moreover, humans are often unreliable as well. Historically, many minority groups have faced unfair treatment in the case of mortgage lending and while since then, many laws have been passed to protect these groups, biases often still exist (either consciously or subconsciously)(Ladd, 1998). It is also much easier to program bias out of machine than humans and with further advancements and more awareness, that is becoming easier than ever(Uzzi, 2020).

Ideally, the best approach would be to integrate both humans and machines when working in this sector. This is called collaborative intelligence and can look like anything- from humans assisting machines(training/retraining, explaining, and sustaining the models/algorithms) to machines assisting humans.

Main Takeaways

In the past, the way the models were trained is now completely outdated. Older data is not representative of an equitable world and hence merely training and testing on historic data without making any changes to our approach is going to lead to more prejudice and an unending cycle of unfairness in the world. As ethical statisticians working in finance, our main goal should be to make our models explainable, transparent and fair by avoiding intentional or unintentional wrongdoings.

Essentially, we should be asking ourselves this question - If everything else is held equal (all non protected attributes), does our model treat protected and non protected classes the same way?

It is also important to remember that the process of being an ethical statistician does not end after developing the model and testing it once. Requirements are constantly changing and the only way we can keep up with AI is by continuous monitoring.

Further reading/ Additional resources:

This is an ever developing field and new resources are available very regularly.

However, a book that will almost always be relevant and would help develop your understanding is Weapons of Math Destruction by Cathy O'Neil. For those who prefer watching/listening instead of reading, a webinar series that I found very helpful is available here

There are also plenty of resources available online that provide great guidelines for being an ethical statistician. Some of these include:
- ASA ethical guidelines
- SSC Code of conduct
- Fair Lending Laws and Regulations

References

- Digalaki, E. (2022, February 2). The impact of artificial intelligence in the Banking Sector & How Ai is being used in 2022. Business Insider. Retrieved April 14, 2022, from https://www.businessinsider.com/ai-in-banking-report
- Explainable AI. IBM. (n.d.). Retrieved April 14, 2022, from https://www.ibm.com/watson/explainable-ai Ladd, H. F. (n.d.). Evidence on discrimination in mortgage lending - jstor.org. jstor. Retrieved April 14, 2022, from https://www.jstor.org/stable/2646961
- Lou, S., & Yang, M. (2020, August 12). Things you need to know before you become a data scientist: A beginner's Guide to Data Ethics. Medium. Retrieved April 14, 2022, from https://medium.com/big-data-at-berkeley/things-you-need-to-know-before-you-become-a-data-scientist-a-beginners-guide-to-data-ethics-8f9aa21af742#:~:text=%E2%80%9CData%20ethics%20is%20a%20new,robots
- Orrell, D., & Houshmand, M. (1AD, January 1). Quantum propensity in economics. Frontiers. Retrieved April 14, 2022, from https://www.frontiersin.org/articles/10.3389/frai.2021.772294/full
- Townson, S. (2020, November 6). Ai Can Make Bank Loans More Fair. Harvard Business Review. Retrieved April 14, 2022, from https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair
- Uzzi, B. (2020, November 4). A simple tactic that could help reduce bias in AI. Harvard Business Review. Retrieved April 14, 2022, from https://hbr.org/2020/11/a-simple-tactic-that-could-help-reduce-bias-in-ai?ab=at_art_art_1x1
- Why every financial institution should consider explainable AI. BNY Mellon. (n.d.). Retrieved April 14, 2022, from https://www.bnymellon.com/apac/en/insights/all-insights/why-every-financial-institution-should-consider-explainable-ai.html