WELLS FARGO

Ethics and Bias in Algorithms

Jie Chen
Joint work with Vijayan N. Nair
Corporate Model Risk, Wells Fargo

June 4, 2020

Fairness/bias in the bank

- Many views of this deck are based on Wells Fargo internal paper "Bias, Fairness, and Accountability with AI/ML Algorithms" by Nengfeng Zhou, Rachel Bartolomeo, Zach Zhang, Vijayan N. Nair, Harsh Singhal, Jie Chen and Agus Sudjianto
- Fairness/bias in the bank
  - Fair lending during the loan life cycle, including marketing, application decisioning, pricing and servicing.
    - The Fair Housing Act (FHA, 1968) and Equal Credit Opportunity Act (ECOA, 2017) prohibit unfair and discriminatory practices based on protected attributes in consumer lending.
    - ECOA explicitly mentions nine categories – race, color, religion, sex, national origin, age, marital status, receipt of public assistance, or exercised any right under the Consumer Credit Protection Act
    - Credit scoring modeling
  - Financial crimes (fraud detection and anti-money laundering).
  - Conduct analysis and compliance management relying on analysis of unstructured data (text, e-mails, etc.) that use natural language processing (NLP) techniques
    - Unfair, Deceptive, or Abusive Acts or Practices ("UDAAP")

Potential Sources of Bias and Discrimination

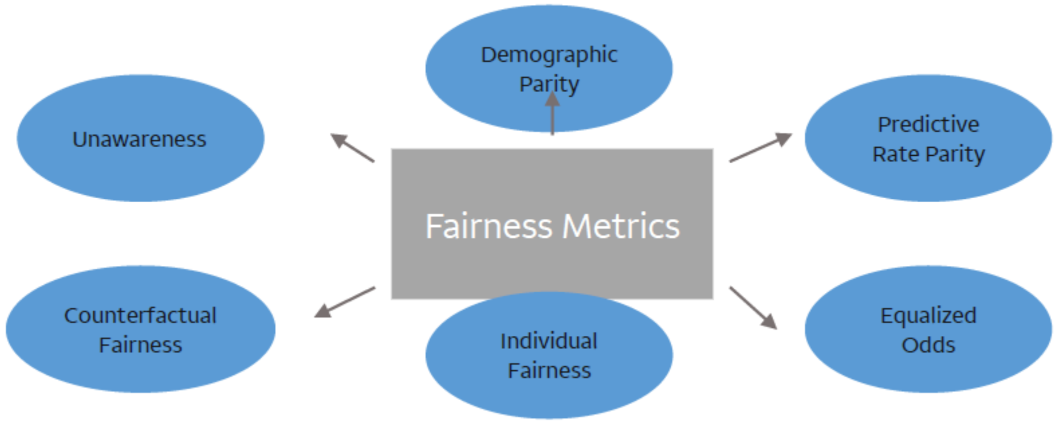
- Data bias
  - Bias in historical data
  - Bias in data collection mechanisms
  - Bias in alternate sources of data
  - Unobservable Outcomes
  - Bias in unstructured data and feature engineering

- Algorithmic bias

  The nature of ML algorithms make the problem more challenging.
  - The automated nature of modern ML algorithms in variable selection and feature engineering
  - The flexible nature of the algorithm—good fitting and memory of patterns
  - The opaqueness of complex ML algorithms

Fairness Metrics

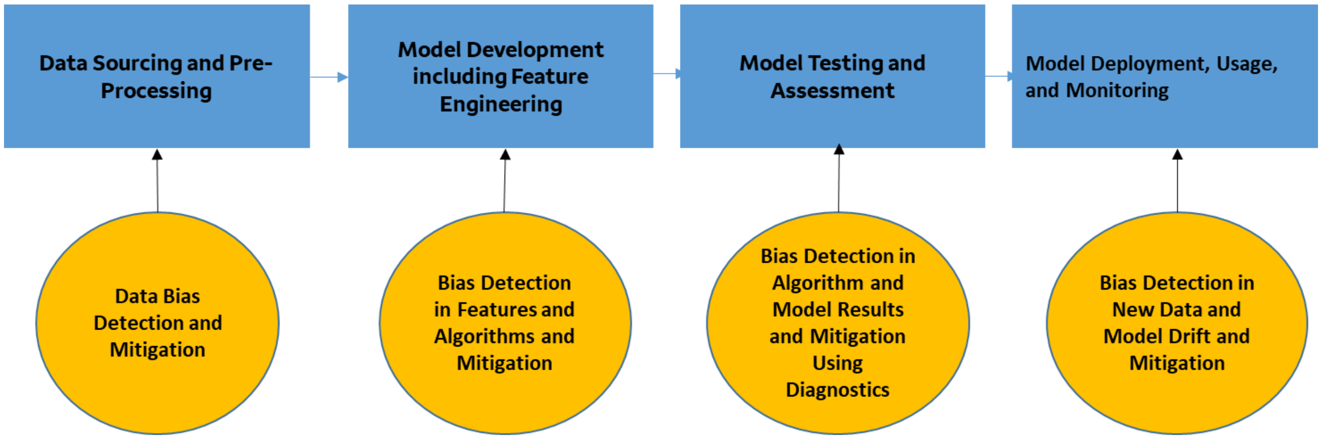
- Common and widely used metrics to assess model or algorithmic fairness.



- Fairness metrics in banking industry
  - There are many regulatory requirements.
  - Fair lending: the statistical analysis typically focused on whether prohibited class (race/gender etc) will play a role in lender's decision for **similarly situated groups**
    - Similarly situated groups: legitimate, non-discriminatory credit characteristics considered by an entity in its underwriting processes, such as FICO and other lending criteria.
  - Conditional parity—demographic parity conditional on business attributes which are accepted as nondiscriminatory.
  - Relationship to other metrics (Equalized odds, individual fairness, causal inference, etc.)

De-biasing and Mitigating Unfairness

- An overview model life cycle, different types of biases and mitigation efforts



- Some of above suggest actively manipulating the data or using protected attributes in the modeling cycle, and their usefulness may be limited due to legal and compliance issues in the bank.
  - To avoid unlawful discrimination under ECOA and Regulation B, lenders generally must not use prohibited basis data or proxies for discrimination in their credit underwriting systems.
  - Separate fair lending group conducts periodic backtesting and trend analysis to validate that credit underwriting systems do not discriminate against applicants on a prohibited basis.

Challenge on privacy data

- Availability of protected attributes.
  - Banks are not legally allowed to collect protected attribute data in lending, except in mortgage loan applications
  - In assessing fair lending, the protected attributes (gender and race) are usually imputed from databases of names and addresses (e.g., Bayesian Improved Surname Geocoding (BISG) method). However, such imputation can be inaccurate, especially for some racial groups.
  - For NLP applications, the information required for the BISG method is not available.