DATA+AI
SUMMIT 2022

Productionizing ethical AI credit-scoring

Jeanne Choo
Specialist Solutions Architect Databricks

ORGANIZED BY databricks

Speaker introduction

Specialist Solutions Architect at Databricks

- Help customers productionize their machine learning applications

Graduated in Plant Sciences and Zoology

- Improve photosynthesis in rice
- Dissect cancer pathways in B-cell lymphoma

Unifying theme

- Career impact

Overview

Why should we care?

Why can AI-driven credit scoring be unfair?

Some existing solutions to mitigate unfairness:

- Specialised metrics eg. equal odds
- Assessment methodologies released by regulators

Remaining challenges

- **Track** fairness metrics across different code, data and model combinations
- **Make auditing simple** by standardizing and automating reporting

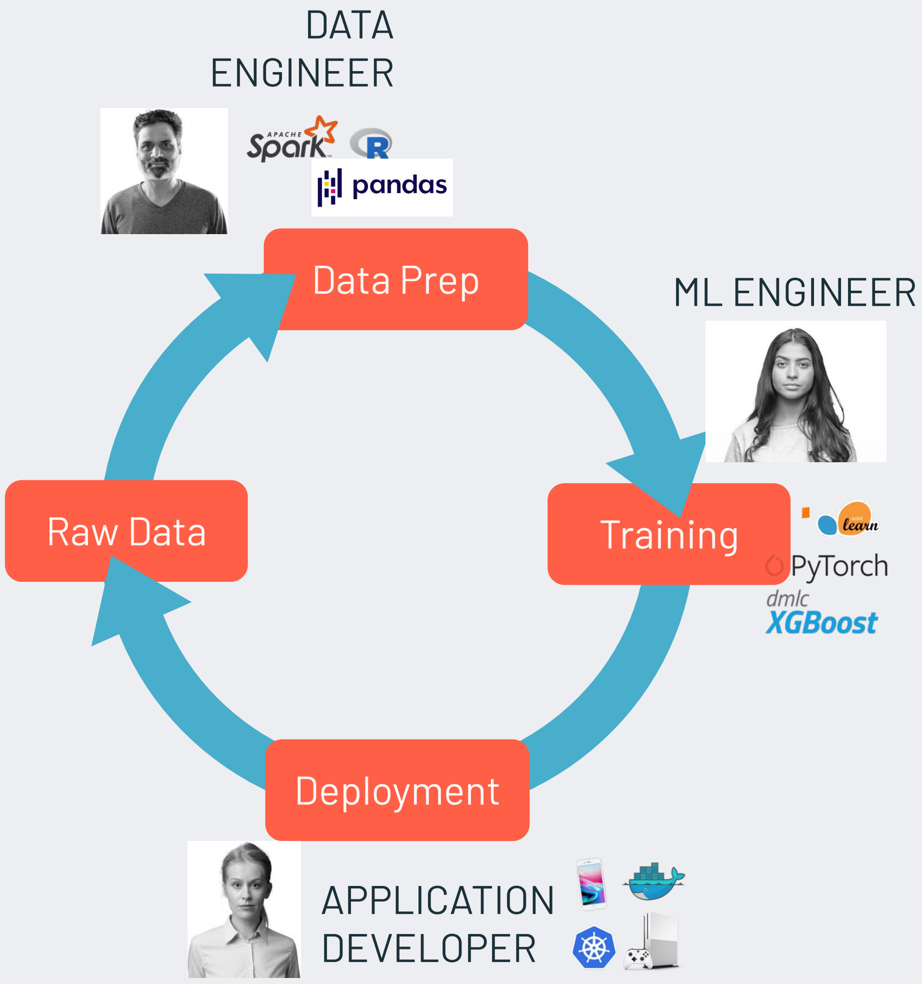


Image credits: Alex Ott, Databricks

Why should we care?

Automated credit scoring systems have high stakes

Determine life outcomes eg. housing loans and employability

But can systematically disadvantage vulnerable groups

Strong need for fairer systems

Why can AI-driven credit scoring systems be unfair?

Historical bias in data

For **historically underserved** groups, credit scores are **noisy indicators** of default risk

Poor quality credit data, not poor model fit, is the main cause of noise

How Costly is Noise?

Data and Disparities in Consumer Credit*

Laura Blattner
Stanford University

Scott Nelson
Chicago Booth

source

Lack of clear definitions of fairness

146 papers analysing "bias" in NLP systems had **"vague"** and **"inconsistent"** motivations that were **"lacking in normative reasoning".** source

Language (Technology) is Power: A Critical Survey of "Bias" in NLP

Su Lin Blodgett
College of Information and Computer Sciences
University of Massachusetts Amherst
blodgett@cs.umass.edu

Solon Barocas
Microsoft Research
Cornell University
solon@microsoft.com

Hal Daumé III
Microsoft Research
University of Maryland
me@hal3.name

Hanna Wallach
Microsoft Research
wallach@microsoft.com

Previously abstract ideas such as "creditworthiness" and "risk-to-society" are now **forcibly quantified**

Operationalisation Challenges

Academic literature has extensive treatment of algorithmic fairness

From the recent Neurips 2021 Conference:

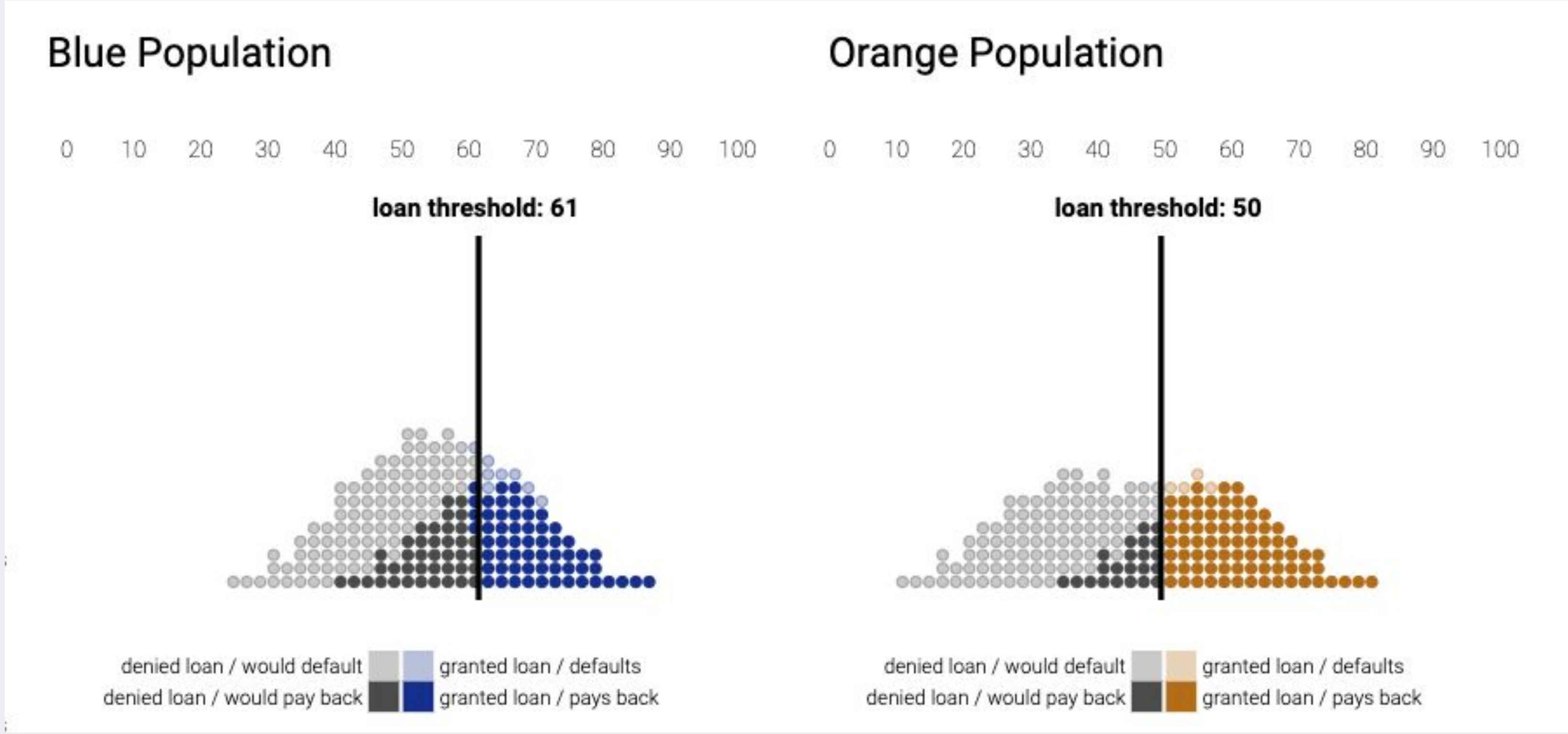
- Algorithmic Fairness through the lens of Causality and Robustness
- On the Impossibility of Fairness-Aware Learning from Corrupted Data
- Fairness for Robust Learning to Rank
- Fair SA: Sensitivity Analysis for Fairness in Face Recognition
- Counterfactual Fairness in Mortgage Lending via Matching and Randomization

Industry players lack guidelines that are practical and actionable.

- Monetary Authority of Singapore Veritas Framework
- Financial Conduct Authority UK – Machine Learning in UK Financial Services

Current Solutions
- Fairness-aware metrics
- Regulatory frameworks

Max profit



Source: http://research.google.com/bigpicture/attacking-discrimination-in-ml/

Different thresholds

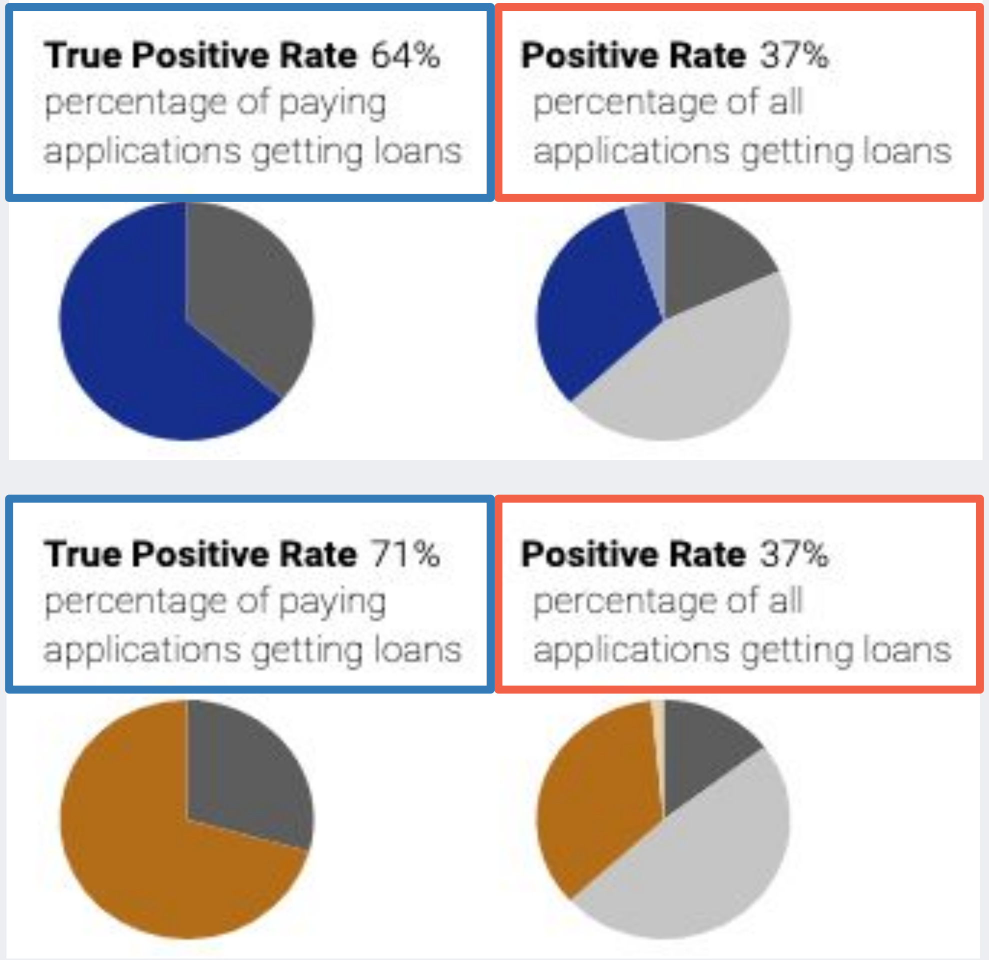
Groups held to different standards

60% paying applicants in blue group correctly granted a loan versus 78% in orange group

Demographic parity

Percentage of granted loans equal between groups

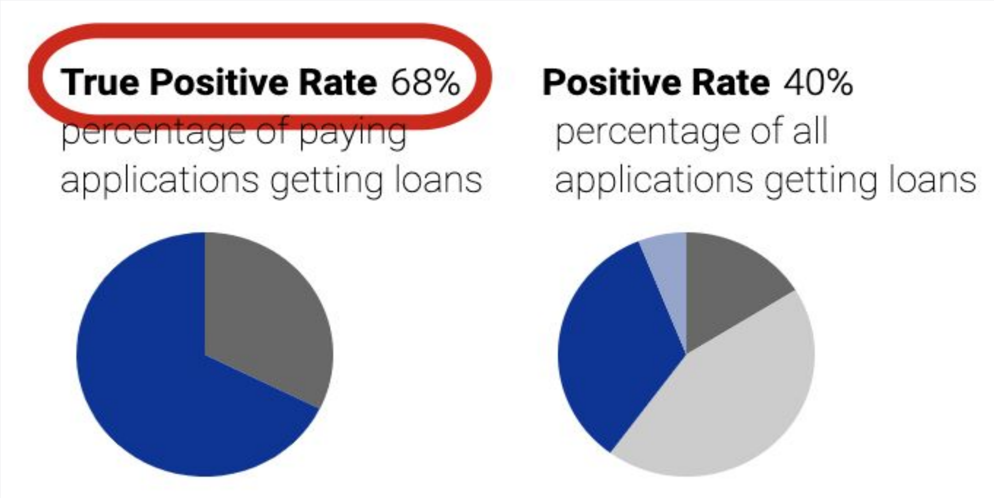
Fewer qualified people in blue group (64%) are granted loans compared to 71% in orange group

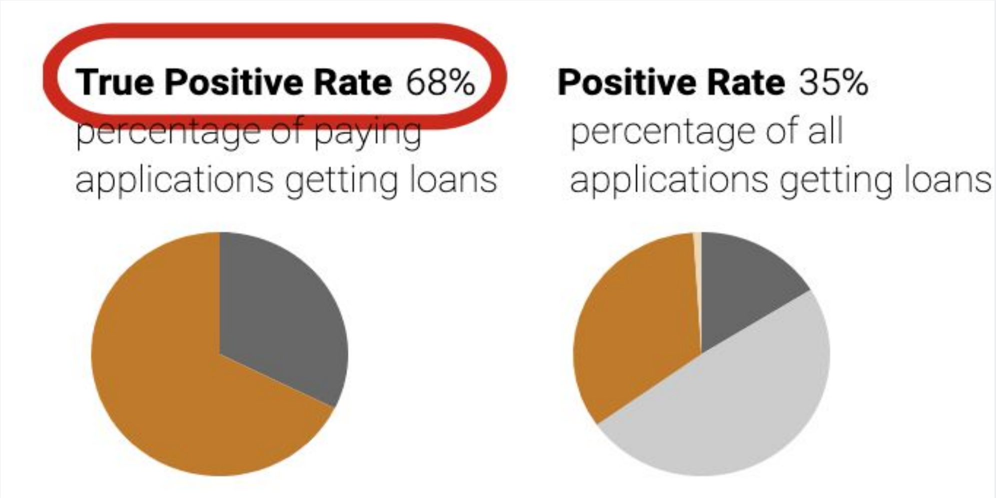


Source: http://research.google.com/bigpicture/attacking-discrimination-in-ml/

Equal opportunity

Among applicants who would pay back loan, percentage of loans between groups are equal



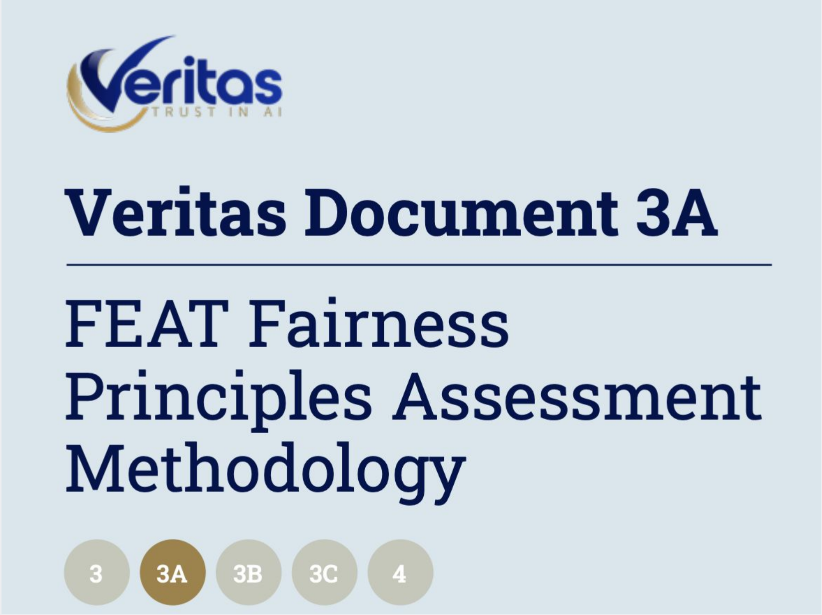


Regulatory frameworks

MAS Veritas Consortium Assessment Methodologies

What is **corrective procedures** are implemented to improve fairness for potentially disadvantaged groups, for example, **post-hoc calibration?**

What **trade-offs** exist between maximising commercial metrics and ensuring fairness? Has an analysis been done?



Remaining challenges

"Enforcing fairness for production-ready ML systems in Fintech requires specific engineering commitments at different stages of ML system life cycle"

-FICO AI Research

Features of a production-ready ethical credit scoring system

Reproducible

Versioned parameters, code and data

Feature engineering logic is consistent between training and serving

Fairness is quantified

Measure system fairness pre-deployment

Protected variables are monitored for production drift

Auditable

Models, even previous ones, are transparent and searchable

Demo

https://tinyurl.com/ethical-ai-dais-2022

Metrics and technology are only a few tools to improve machine learning

Ensuring fair systems will involve a multi-disciplinary approach

References

Open source packages

Monetary Authority of Singapore veritastool Github

Definitions of Fairness

Counterfactual Fairness

On Fairness and Calibration

How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness

Fairness and Machine Learning

Hands-on Fairness tutorial

Interactive visualisation of different fairness metrics

Causes of ML bias

How Costly is Noise? Data and Disparities in Consumer Credit

Language (Technology) is Power: A Critical Survey of “Bias” in NLP