THE PPM FRAMEWORK: A NEW WAY OF AUDITING ALGORITHMIC SYSTEMS
A Conceptual Introduction to Machine Learning

Kareem Saleh, Founder and CEO of FairPlay
January 19, 2022
What is Machine Learning?

Traditional Programming

Ingredients + Recipe = Outcome
But what if you don’t have a recipe?

Traditional / Programming

Ingredients + Recipe = Final Product

Machine Learning Algorithm

Ingredients + Outcome = Recipe

MANAGING SPACE
How is Machine Learning Being Used in Housing?

- **TENANT SCREENING**
- **MORTGAGE UNDERWRITING**
- **HOME APPRAISAL**
Lending decisions used to be made by humans who tried to assess an applicant’s creditworthiness.
In the 1980s we started using math instead of humans to make lending decisions because it was seemingly more “neutral” and “objective”
The old math assumes relationships between ingredients (variables) are straight-forward...
The old math assumes relationships between ingredients (variables) are straight-forward, but the world is actually complex and nonlinear.
Imagine we tried to build a model that predicts sex

**CAN WE DETERMINE GENDER USING HEIGHT?**

+ Men, on average, are taller than women.
Height is somewhat but not perfectly predictive of sex

CAN WE DETERMINE SEX USING HEIGHT?

+ Men, on average, are taller than women.
+ But there are tall women and short men, and many people that are the same height.
+ Using height alone isn’t very accurate.
Including weight adds predictive power, but the model still isn’t perfect

CAN WE DETERMINE SEX USING HEIGHT + WEIGHT?

Because men are, on average, heavier than women of the same height, accuracy would improve.
Using height and weight to predict gender causes kids to be classified as women

CAN WE DETERMINE SEX USING HEIGHT + WEIGHT?

+ Because men are, on average, heavier than women of the same height, accuracy would improve.

+ *But children would mostly be misclassified as women.*
Is birthdate predictive of Sex?

CAN WE DETERMINE SEX USING HEIGHT + WEIGHT + BIRTHDATE?

Knowing people’s age eliminates the misclassification of children, and improves the model’s accuracy of determining gender across all age groups.

+ **But the idea that birthdate can help determine sex is not obvious**
When there are a lot variables, the recipe connecting them to an outcome can be so complex no one can understand it:
When there are a lot of variables, the recipe connecting them to an outcome can be so complex that no one can understand it: a black box.
The problem with the black box is the risk it will be biased in ways we don’t understand.
If you're buying a high mileage car in Nevada there’s a big probability you’re a person of color
ML algorithms relentlessly refine their “recipe” to achieve the best outcome (aka they adjust to better accomplish their target or objective)

Every algorithm must be given a target
Social media algorithms seek to maximize their target: engagement

Every algorithm must be given a target

Social media target: Maximize engagement
The algorithm single-mindedly focuses on engagement regardless of whether it’s good for your health or good for society.

Every algorithm must be given a target.

Social media target: Maximize engagement.

Without regard for societal harm.
Giving an algorithm one target is problematic: imagine a self-driving car whose only target was to get you from point (a) to point (b)

Target: Get from point (a) to point (b)
Self-driving cars have a second target: Safety

Target: Get from point (a) to point (b)

Second Target: Safety (obey traffic laws; avoid accidents with cars, pedestrians, cyclists, etc)
We can do this in financial services: Target a low risk of default...
We can do this in financial services: Target a low risk of default while also targeting fairness.

Target: Low risk of default

Second Target: Fairness
U.S. Mortgage Fairness in 2020: Female

ADVERSE IMPACT RATIO:

- Less than 80%
- 80 to 90%
- Over 90%
U.S. Mortgage Fairness in 2020: Black

ADVERSE IMPACT RATIO:

- Less than 80%
- 80 to 90%
- Over 90%
A fairness-optimized neural network can produce very accurate and considerably fairer mortgage outcomes than a logistic regression or unconstrained neural network.
A fairness-optimized neural network can produce very accurate and considerably fairer mortgage outcomes than a logistic regression or unconstrained neural network.
A fairness-optimized neural network can produce fairer mortgage pricing than a linear regression or unconstrained neural network.

SMD values less than 0.2 are generally considered acceptable; all of the models shown here are well within that range.
The fairness constrained neural network generates fairer results using the same variables.

### Adverse Impact Ratio

<table>
<thead>
<tr>
<th>Model</th>
<th>Adverse Impact Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>90%</td>
</tr>
<tr>
<td>Unconstrained Neural Network</td>
<td>90% 56%</td>
</tr>
<tr>
<td>Fair Neural Network</td>
<td>90% 96%</td>
</tr>
</tbody>
</table>

### Ranking

<table>
<thead>
<tr>
<th>Rank</th>
<th>Reason</th>
<th>Fairness Constrained Neural Network</th>
<th>Unconstrained Neural Network</th>
<th>Fair Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Borrower’s score Previous December</td>
<td>60%</td>
<td>56%</td>
<td>96%</td>
</tr>
<tr>
<td>2</td>
<td>Borrower’s score Previous September</td>
<td></td>
<td>56%</td>
<td>96%</td>
</tr>
<tr>
<td>3</td>
<td>Borrower’s score Previous June</td>
<td></td>
<td>56%</td>
<td>96%</td>
</tr>
<tr>
<td>4</td>
<td>Borrower’s score Previous March</td>
<td></td>
<td>56%</td>
<td>96%</td>
</tr>
<tr>
<td>5</td>
<td>Is this the borrower’s first mortgage?</td>
<td></td>
<td>56%</td>
<td>96%</td>
</tr>
</tbody>
</table>

**Reasons:**
- **Borrower’s score**: Previous December
- **Borrower’s score**: Previous September
- **Borrower’s score**: Previous June
- **Borrower’s score**: Previous March
- **Co-borrower’s score**: Previous December
- **Co-borrower’s score**: At origination
- **Loan-to-value**
Catalysts driving AI’s takeover of high stakes decisions

- Data
- Advanced Algorithms
- Cloud Computing
Algorithmic auditing – why and how

Alex C. Engler
Twitter: @aengler
Email: aengler@brookings.edu
Proliferation of algorithms in important services

• “Over 75% of colleges and universities use analytics for enrollment management” – EduCause Survey, 2015

• “55% of human resources leaders in the United States use predictive algorithms in hiring, at many stages” – Mercer Global Talent Trends

• “98 percent of mortgages originated by Quicken Loans, the country’s largest lender, used the company’s digital platform, Rocket Mortgage” writes the NYTimes

• “We found that a category of algorithms that influences health care decisions for over a hundred million Americans shows significant racial bias,” said Sendhil Mullainathan
Characteristics of algorithms in important services

- Scaled – they often affect many people
- Automated – they often replace an existing process that was done by a person (but not always)
- “Black box” – They may obscure the decision-making process & reduce efficacy of civil liability
- Might be consistent in treatment (but often not in impact)
- “Learning” means algorithmic behavior may change over time
Concerns with algorithms in important services

- Power centralizing towards institutions (e.g. tuition, ride-sharing)
- Incentivizes privacy violations (e.g. Clearview AI, EverAlbum)
- Opaque process may offer no feedback (e.g. mortgage denial)
- Fraud (different than efficacy)
- Discrimination
- Systemic Risks (e.g. flash crashes)
Proposed solution: Algorithmic audits!

- An algorithmic audit is a review of the environment, inputs, functions, outputs, and results of an algorithmic system.

- This is what the NFHA’s Purpose, Process, Monitoring Framework describes, but there are many other examples from health, hiring, education, social media, etc.

- Can be paired with a fine for illegal activity, possibly algorithmic deletion in extreme cases.
Do algorithmic audits help with these problems?

- Power centralizing towards institutions (e.g. tuition, ride-sharing)
  - Mostly no, may create public awareness
- Incentivizes privacy violations (e.g. Clearview AI, EverAlbum)
  - Yes, if the privacy violations are illegal
- Opaque process may offer no feedback (e.g. mortgage denial)
  - Mostly no – audits offer global, not individual, explanations.
- Fraud
  - Yes!
- Discrimination
  - Yes!
- Systemic Risks
  - Maybe
What should an audit be?

**Process considerations:**
- Independent
- Adversarial
- Representative (if there are many models)
- Access to data/code/models

**Model & outcome review**
- Look at outcome/training/hypothetical data
- Consider problem definition
- Consider algorithmic drift, frequency of audit

**Dependencies & Docs**
- Data collection process
- Training data representativeness
- Dependency analysis
- Accurate documentation
Who needs to do audits?

• Criteria
  • Audits must be **independent** and **adversarial**.
  • The require extensive access to the data and algorithms
  • This will often mean government action

• Audits do not automatically mean accountability. In accounting, it takes:
  • Market incentives (these don’t apply to algorithmic discrimination)
  • Financial auditors (the people!) are legally liable if they commit fraud
  • Plus direct government oversight (SEC)
So, Government Audits

• Why:
  • They can be genuinely independent.
  • Usefully, fraud & discrimination are often already against the law.

• How:
  • Develop specific best practices/guidance by model/application type
  • Some legal changes may be necessary
    • Expand administrative subpoenas to include data sets
    • E.g. remove ‘predictive validity’ from EEOC Uniform Guidelines
  • Hire regulatory data scientists (soon to be a booming field)
  • Start information gathering, enable consumer complaints, and encourage industry whistleblowing
Government Agency Action

- HHS was only agency to detail its algorithmic regulatory authority
- EEOC launched Initiative on AI Fairness
- Federal Trade Commission
  - Threatening blog post on fraud and discrimination
  - Fines and algorithmic deletions for privacy violations (EverAlbum)
  - Started rulemaking consideration

- Five financial regulators started information gathering on algorithms, citing Fair Housing + Equal Credit Opportunity Acy
  - (DOT OCC, Fed Reserve, FDIC, CFPB, NCUA)
- Housing and Urban Development
  - Reversing the Trump admin change to the ‘discriminatory effects’ rule
  - Has not responded to the two E.O.s on algorithms (E.O. 13960 and E.O. 13859)
Aside: Agencies have algorithms too

- Federal agencies can show the way by documenting and auditing their own algorithms
- Algorithmic documentation required under E.O. 13960
- Eventually, standards for algorithmic auditing will be the norm
Complications of algorithms in important services

• Many algorithmic stages of “funnel processes;” consider hiring:
  • Job recommendation -> Resume analysis -> Questionaries/Games -> Automated interviews

• Different types of algorithms combined, consider college tuition:
  • First: predict what one student would pay to attend
  • Second: assign scholarships across many prospective students

• Separate algorithmic systems affecting one another (e.g. online markets)

• Human + algorithmic decisions can have unexpected outcomes
  • Felony sentencing by judge with algorithmic risk assessment
Work supported by:

STIFTUNG MERCATOR  

FULBRIGHT Schuman Program

Get in touch:
Twitter: @alexcengler
aengler@brookings.edu
PPM AUDITING FRAMEWORK

A COMPREHENSIVE NEW FRAMEWORK FOR AUDITING ALGORITHMIC SYSTEMS

AI AUDITING IN HOUSING

Lisa Rice, President and CEO of NFHA
MARCH 22, 2022

© 2022 National Fair Housing Alliance. Please do not distribute.
Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women.

Healthcare

Credit Scoring

Employment

Racial bias in a medical algorithm favors white patients over sicker black patients.
Apple’s credit card is being investigated for discriminating against women

Credit

Tenant Screening Services
TECH IS THE NEW CIVIL RIGHTS FRONTIER

• Credit Scoring Systems
• Automated Underwriting Systems
  • Automated Verifications Processing
• Risk-based Pricing Systems
• Advertising Systems/Platforms
• Revenue Management Software
• E-loans Processing
  • Video-Conferencing
  • Digital Fingerprinting
• Facial Recognition Systems
• Automated Property Valuation Systems
• Tenant Selection Systems
Bifurcated U.S. Financial System

Capital Markets
- Banks, Credit Unions, Savings and Loan Companies, CDFIs, GSEs, Federal Home Loan Bank Boards, Mutual Funds, Pensions, 401(k)s, Stocks, Bonds, AAA Rated Mortgage-Backed Securities

Mainstream Financial Services
- Prime Mortgages, Savings and Checking Accounts, Home Equity Loans, Lines of Credit, Certificates of Deposit, Prime Auto Loans
- Prime Market

Fringe Financial Services
- Pawnshops, Check Cashers, Payday Lenders, Rent-to-Own Shops, Title Lenders, Finance Lenders, Sub-Prime Lenders, Buy Here Pay Here Auto Lenders

Middle / Upper-Income and Predominately White Communities

Source: Carr, Jim, Lisa Rice and Shanti Abedin
• Unrepresentative, Insufficient or Flawed Data
• Design Flaws
• Biased Feedback Loops
• Insufficient or No Testing for Bias
• Untrained Designers
• Lack of Diversity
Compelling models to be fairer can lead to the development of more accurate and safer models that minimize or eliminate disparities in housing and lending.

The adoption of fairer models would expand access to credit for underserved groups and result in less disruption for consumers, families, communities, and industry.

Fairer models can help reduce the racial wealth and homeownership gaps.

The development of state-of-the-art policies and protocols would mark the U.S. as a leader in advancing responsible technologies that serve markets, consumers, and the greater society.
Overview and Purpose Element of the PPM Framework

Snigdha Sharma, FairOps Team Lead & Tech Equity Analyst, NFHA
• What job does PPM – Purpose, Process, Monitoring - do?
• The Purpose element of the PPM Framework?
• Examines every stage of algorithm solution development

• A trust tool regulators can use to gather information that could be used to decide how fair, transparent and explainable an algorithmic system is.

• Helps policymakers see what aspects of algorithmic systems need mitigation and can lead to algorithmic risks to consumers

• Helps policymakers hold those developing algorithms accountable for any decisions they made while building the system
Business Problem: Would this person meet their rent obligation or not?

Input: application information, employment, bank statements etc. of previous or current tenants

Black box: a classification algorithm like Random Forest

Output: tenant score that rates an applicant and is used to decide if an individual is given a lease or not
Business Understanding:

- The Purpose section asks why the model is being built and what are the business objectives.

- Auditors are guided to seek information to make informed decisions about risks the business problem may pose to consumers, institutions, and society at large.

- Does the system pose too great a harm that it should not be built or put into production?

- What exactly is the business problem being solved?
Data Understanding:

• Is there data to represent the business problem?

• Were any techniques used to mitigate risks associated with data paucity or data quality.
Data Understanding:

- Is there proxy testing for protected classes?
- What metrics are being used to test model accuracy and fairness?
- Are model features or variables representative across protected class data?
NFHA PPM Auditing Framework: Process

John Merrill, Chief Technology Officer of FairPlay

March 22, 2022
The Process Stage of the PPM Framework Includes Five Elements

- Staff Profile
- Data Assessment
- Outcome Assessment
- Model Assessment
- Model Use and Limitations
**Staff Profile**

**Diverse Team**
Do the members represent a broad class of possible stakeholders?

**Functional Team**
Can the members work together?

**Communication**
Can the team explain its results to others outside of the team?
**Data Assessment**

- **Collection**
  How was data gathered?

- **Accurate**
  Are the values in the data correct and consistent?

- **Appropriate**
  Is the data well-suited to the model’s objective?

- **Well-behaved**
  Are the data targets and values well-behaved? Are any computed features reasonable?

- **Representative**
  Does the data reflect the characteristics and qualities of the subject being modeled?
Model Assessment

**Model Type**
Is the model type appropriate to the data and problem?

**Parameters**
Which variables are included in the model? How is each one treated?

**Hyper-Parameters**
How did you set the parameters? For example, did you control how fast the algorithm learned? When did you cut off its learning?

**Fairness Constraints & Less Discriminatory Alternatives**
What steps did you take to ensure model was fair? If a model exhibits disparities for a group, did you search for fairer alternatives? If so, how?
Outcomes Assessment: Did the model meet its objectives?

- **Accurate**: Are the model’s predictions correct?
- **Fair**: Do the model outcomes favor one group over another?
- **Explainable**: Are the model’s decisions known and sensible?
- **Stable & Robust**: Do the model’s predictions vary widely when things change?
Model Use and Limitations

Limitations and Assumptions
What does the model assume about the data?

Other Applications
To what other problems can the model outputs be or not be applied?

Disaster Response
Are there procedures and processes in place to handle disasters?
CONTACT

john@fairplay.ai
The Monitoring Element of the PPM Framework

Michael Akinwumi, Chief Tech Equity Officer, NFHA
• Model Use and Model Limitations

• An algorithmic solution needs monitoring

• Q & A (All)
Model Use and Limitations

• Developer needs to be transparent about blind spots in their model. Every model has a limitation.

• What business problem can the model be applied to?
An algorithmic solution needs monitoring

• “All models are wrong, but some are useful.”

• Document what metrics developers plan to use for monitoring
  • The usefulness of the model and
  • Potential harms it may cause to consumers while in production

• Integrity of the model should be protected, and consumer privacy should be protected in the event of a hack

© 2020 National Fair Housing Alliance. Please do not distribute.
Question & Answer