Data Science For Social Good
Carnegie Mellon University

ML
MACHINE LEARNING
DEPARTMENT

HeinzCollege
INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

On Responsible AI/ML/DS, Post-modeling

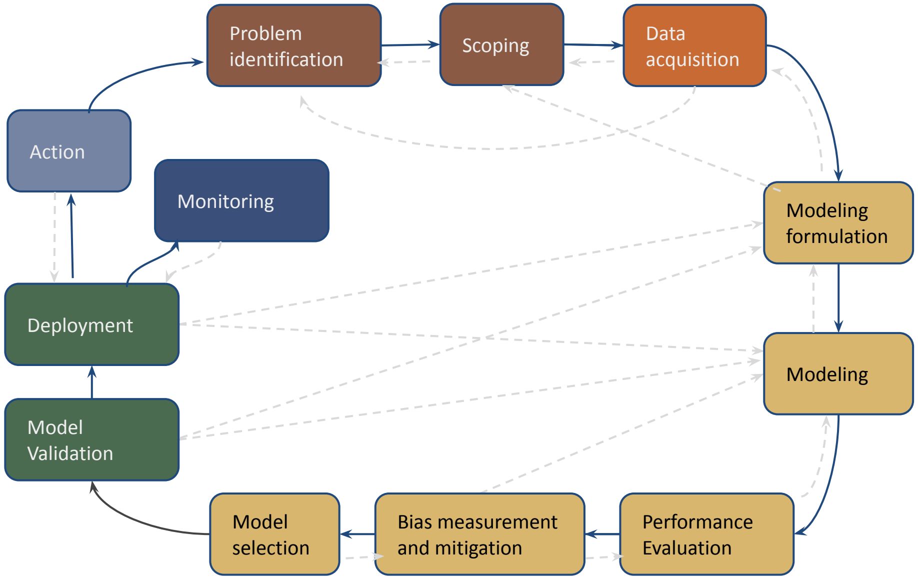
June 2025

Carnegie Mellon University

MACHINE LEARNING DEPARTMENT

Heinz College INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

AI/ML/DS project life cycle



Problem identification

Scoping

Data acquisition

Modeling formulation

Modeling

Performance Evaluation

Bias measurement and mitigation

Model selection

Model Validation

Deployment

Monitoring

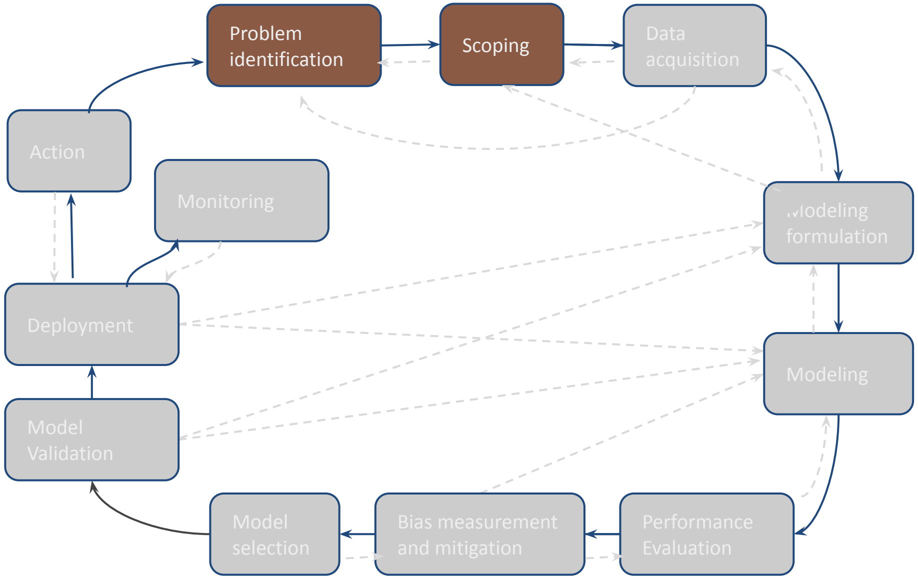
Action

🟫 Ideation and scoping

🟧 🟨 Data curation and modeling

🟩 🟦 Validation, deployment, and monitoring

AI/ML/DS project life cycle



Problem identification

Scoping

Data acquisition

Modeling formulation

Modeling

Performance Evaluation

Bias measurement and mitigation

Model selection

Model Validation

Deployment

Monitoring

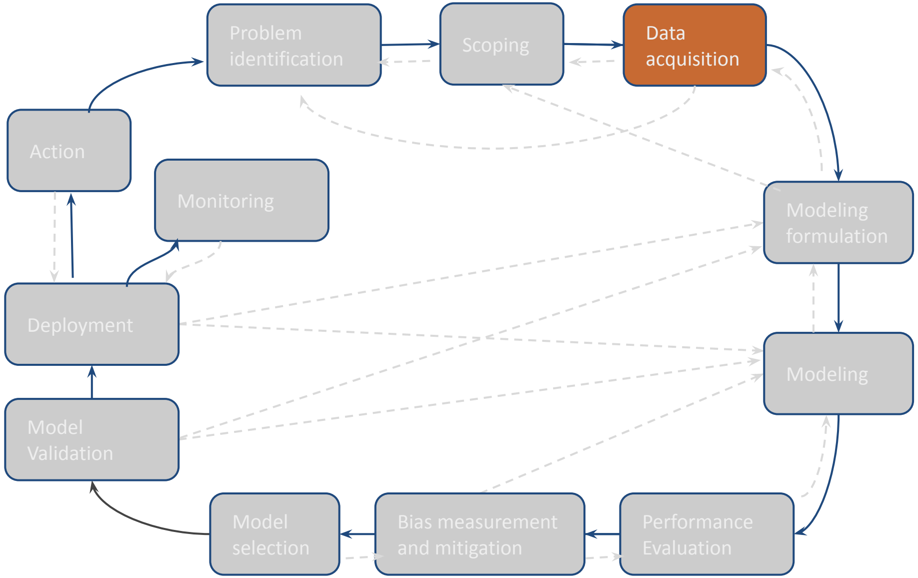
Action

🟫 Ideation and scoping

🟧🟨 Data curation and modeling

🟩🟦 Validation, deployment, and monitoring

AI/ML/DS project life cycle

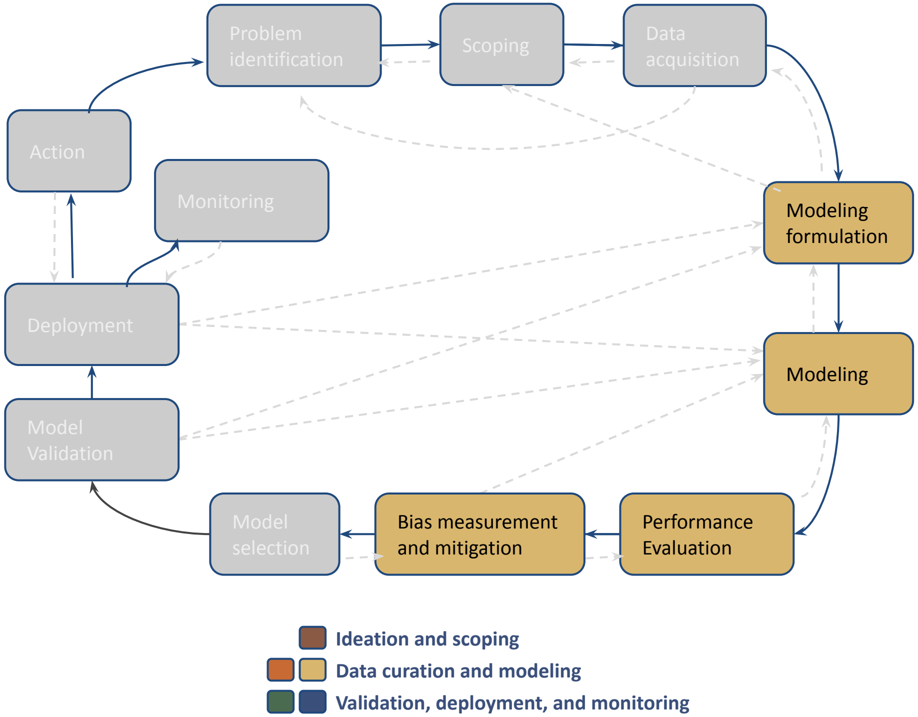


🟫 Ideation and scoping

🟧 🟨 Data curation and modeling

🟩 🟦 Validation, deployment, and monitoring

AI/ML/DS project life cycle



Problem identification

Scoping

Data acquisition

Modeling formulation

Modeling

Performance Evaluation

Bias measurement and mitigation

Model selection

Model Validation

Deployment

Monitoring

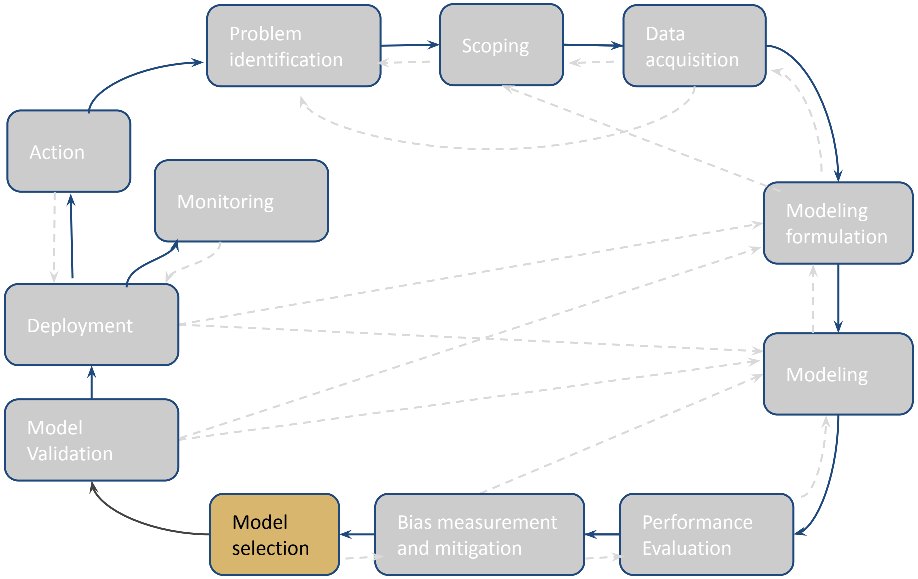
Action

🟫 Ideation and scoping

🟧 🟨 Data curation and modeling

🟩 🟦 Validation, deployment, and monitoring

AI/ML/DS project life cycle

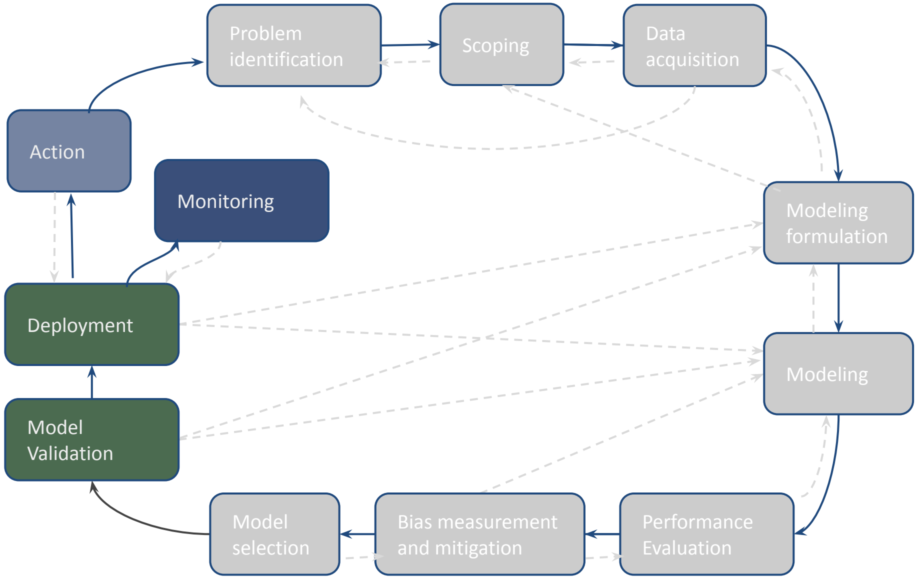


🟫 Ideation and scoping

🟧 🟨 Data curation and modeling

🟩 🟦 Validation, deployment, and monitoring

AI/ML/DS project life cycle



Problem identification

Scoping

Data acquisition

Modeling formulation

Modeling

Performance Evaluation

Bias measurement and mitigation

Model selection

Model Validation

Deployment

Monitoring

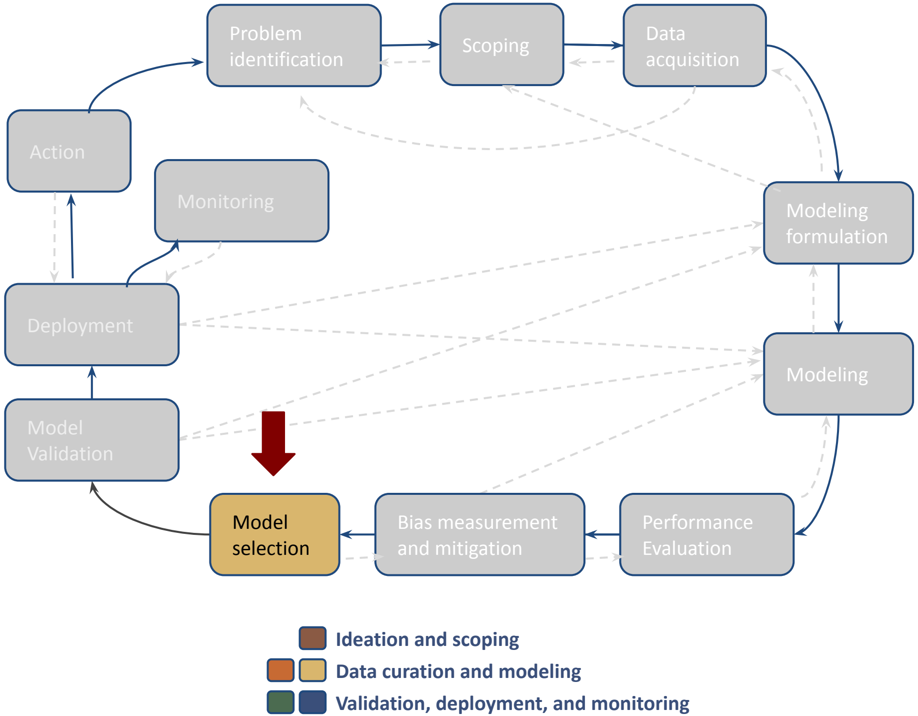
Action

🟫 **Ideation and scoping**

🟧 🟨 **Data curation and modeling**

🟩 🟦 **Validation, deployment, and monitoring**

AI/ML/DS project life cycle

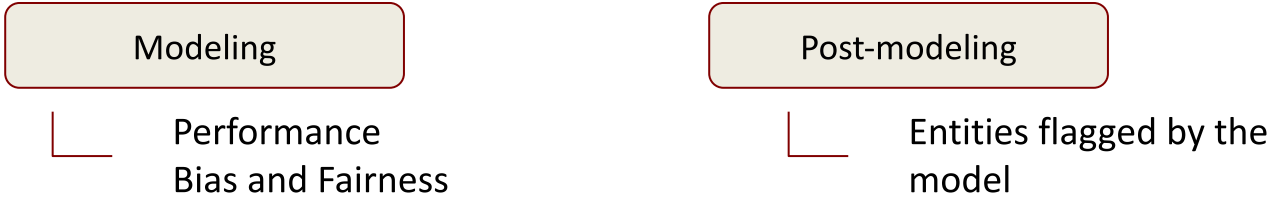


What is Post-modeling?

Deep analysis on a subset of models that best fit the project’s goals –efficiency, effectiveness, equity–

Why

- We need to select “the best” model to deploy with the best possible outcomes for the people it will affect/serve
- This analysis will generate information about the entities the different models in the subset highlight/flag/identify



Modeling

Performance
Bias and Fairness

Post-modeling

Entities flagged by the model

Types of analysis in Post-modeling

Entity id	Score	Label
34	0.765	0
102	0.653	0
765	0.632	1
7	0.517	1
…	…	…
45	0.039	0

Top k

Cohort

Entities with highest likelihood of… according to the model

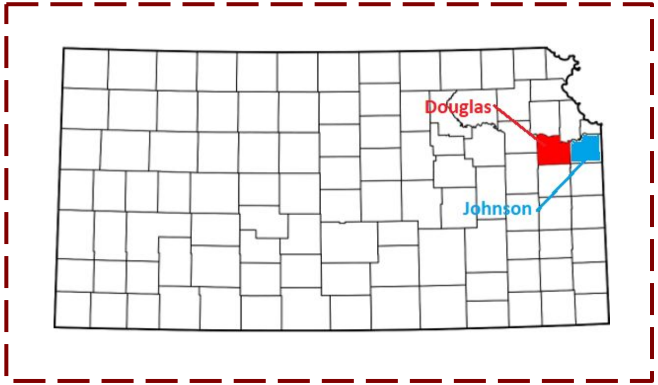
Entities with lowest likelihood of… according to the model

k organization’s capacity to do the intervention

Post-modeling analysis (top k entities)

Type of analysis	What information we get	Comparison level
Crosstabs	Differences in **feature values** between top k selected by model and the rest of the entities.	Single model Between models
Overlaps	Which **entities** are highlighted on different models	Between models
List characteristics	Descriptives (demographics and others) Which **entities** are included Which **entities** are left behind	Single model Between models
Events and outcomes	On label window, After label window	Single model Between models
Error analysis	Which features are associated with FPs, FNs	Single model Between models
Performance	Performance of the models (Precision, Recall, etc.)	Single model Between models
Feature importances	Which **features** add more information to the model	Between models
Bias and Fairness	Group **disparities** at attributes of interest	Between models

Use case: Reducing the impact of Behavioral Health Crises in Douglas and Johnson Counties, Kansas.



Goals

Efficiency

Outreach resources are only allocated to people at-risk of an event

Efficient use of intervention resources

Effectiveness

People selected for outreach are positively impacted by the intervention

Reduced risk of adverse event

Equity

Individuals from high need groups are not left out disproportionately

Fair and equitable distribution of services

Analytical formulation - Matching the operational setup

How often?

Who?

How many?

What outcome are you predicting?

For what purpose?

On the 1st of every month, for all individuals who have interacted with MyRC source agencies in the last 1 year, can we identify the 100 individuals who are at highest risk of having a **very high-acuity*** event in the **next 6 months** to recommend for proactive behavioral health outreach?

*Death by suicide or overdose, suicide atttempts, suicidal gestures, diagnoses, and ambulance runs, overdose ambulance runs, severe substance use, and homicidal intentions or actions.

Current model (ML)

1

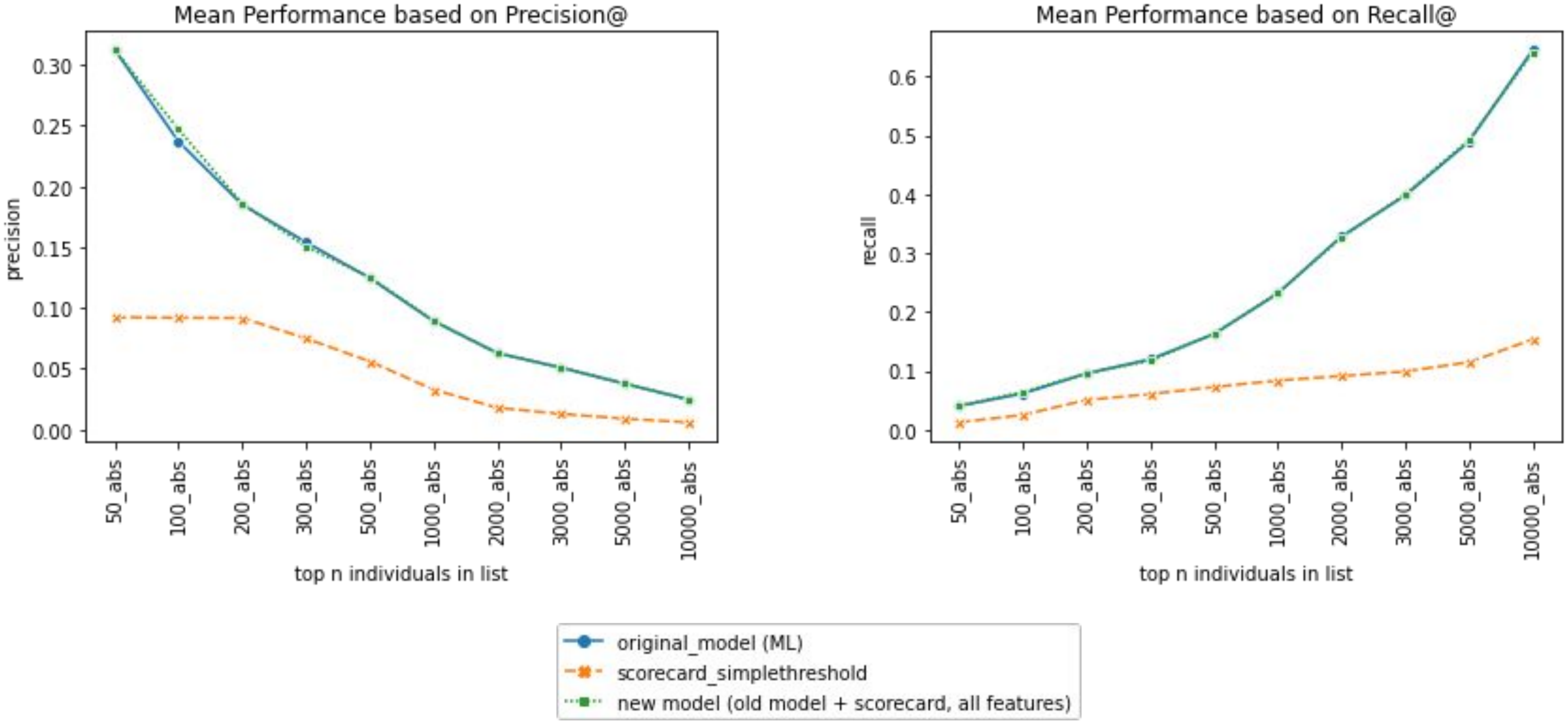
Client Risk Scorecard

2

Information gathered with Post-modeling analysis

Post-modeling analysis	Current model	Scorecard
Demographics (TPs):	● Avg age of 38 ● Same distribution in gender, 49% female ● More in “Other” race than Scorecard	● Avg age of 31 ● Same distribution in gender, 49% female ● More in “Black” race than Current model
Events from past	More events from all types and both acuities	Less events from all types and both acuities
Events on label window	More events from all types and both acuities	Less events from all types and both acuities
Events after label window	More events from all types and both acuities, including deaths	Less events from all types and both acuities, also found deaths
Crosstabs	People at the top have more frequent and higher acuity events	People at the top have more flags on for events considered of high risk
Overlaps	● 0% if only TPs ● Avg of 11% in top 100	
Performance	Better in precision and recall –efficiency, effectiveness–	Less precision and recall
Bias and fairness	Less bias on both attributes of interest (Fair for race)	Unfair for race and gender

Post-modeling - Performance



Crosstabs (last time split `2023-07-01`)

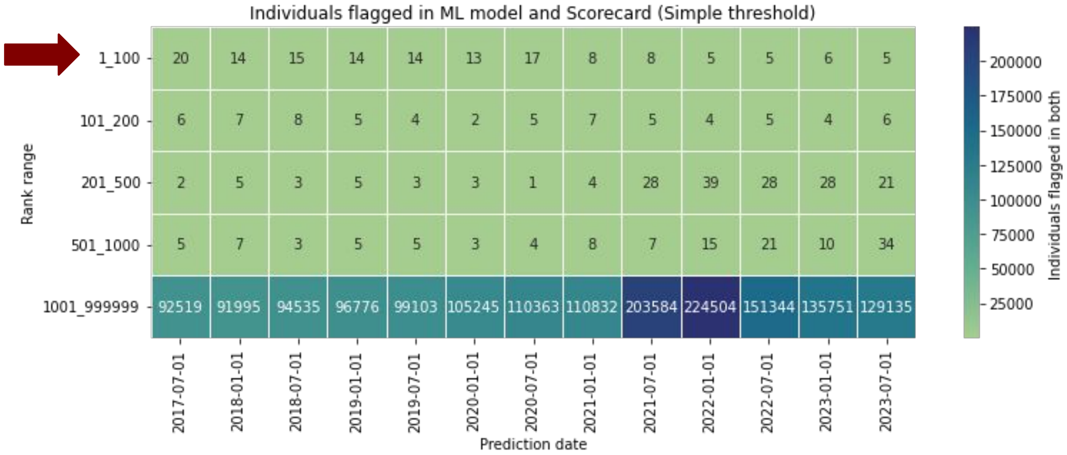
Current model

- **1308** times more likely to have ambulance runs related to homelessness in the last month
- **1308** fold increase in number of ambulance runs related to homelessness in the last month
- **760** fold increase in # of crisis calls (JCMHC) in the last 6 months
- **754** fold increase in # of crisis calls (JCMHC) in the last month
- **739** fold increase in # of ambulance runs related to suicide

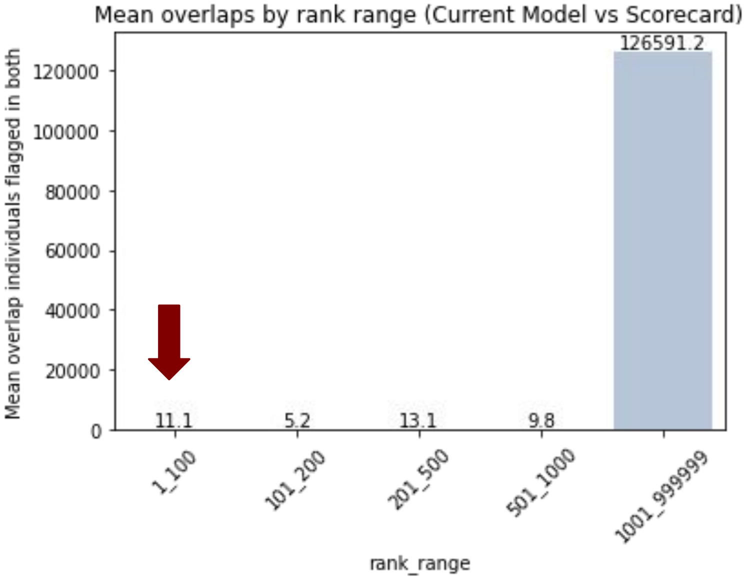
Scorecard

- **311** times more likely to have high risk of substance use
- **229** times more likely to be flagged as High risk of harm to others
- **179** times more likely to be flagged as High risk of suicide
- **145** times more likely to be flagged as High risk of hospitalization
- **88** times more likely to be flagged as High risk of self harm

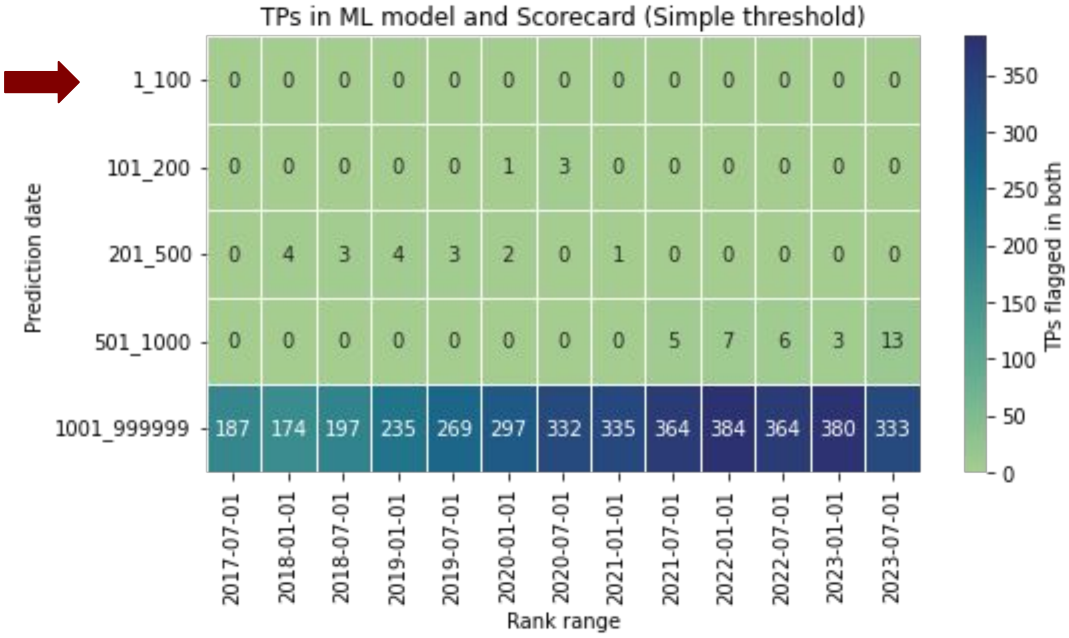
Overlaps between Scorecard and Current model



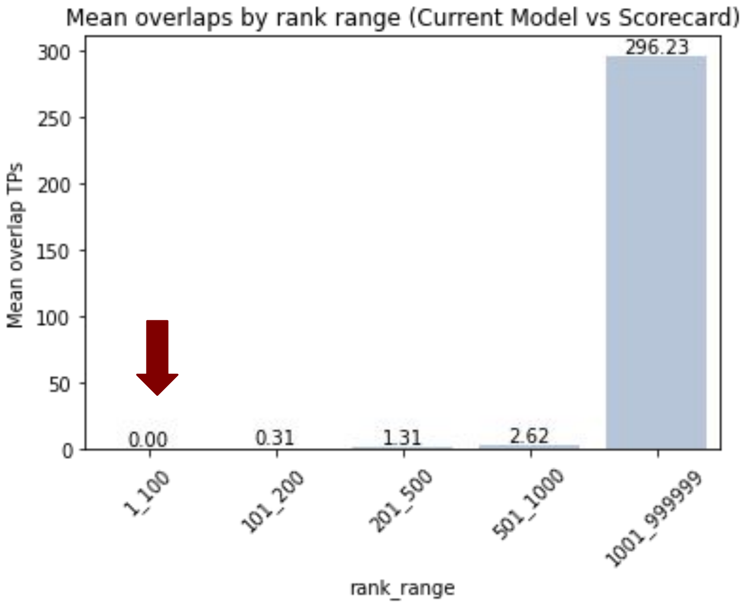
Rank range	2017-07-01	2018-01-01	2018-07-01	2019-01-01	2019-07-01	2020-01-01	2020-07-01	2021-01-01	2021-07-01	2022-01-01	2022-07-01	2023-01-01	2023-07-01
1_100	20	14	15	14	14	13	17	8	8	5	5	6	5
101_200	6	7	8	5	4	2	5	7	5	4	5	4	6
201_500	2	5	3	5	3	3	1	4	28	39	28	28	21
501_1000	5	7	3	5	5	3	4	8	7	15	21	10	34
1001_999999	92519	91995	94535	96776	99103	105245	110363	110832	203584	224504	151344	135751	129135



Overlaps (TPs) between Scorecard (simple threshold) and Current Model

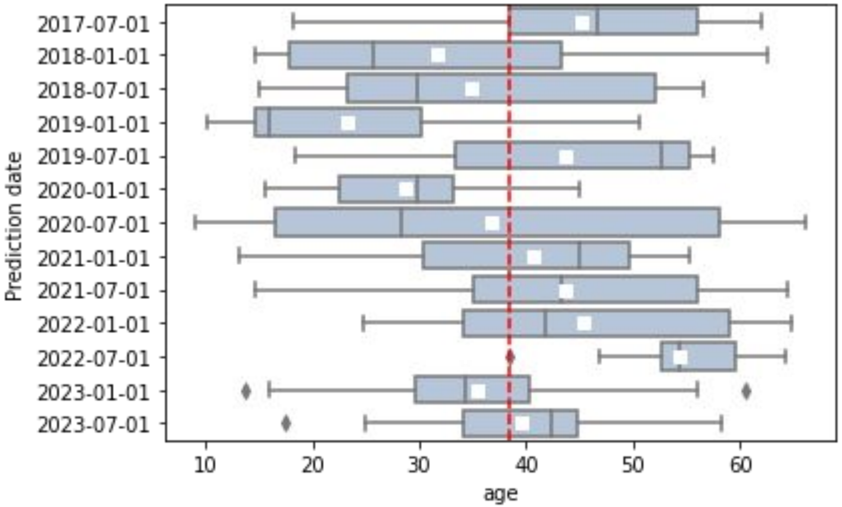


Prediction date	2017-07-01	2018-01-01	2018-07-01	2019-01-01	2019-07-01	2020-01-01	2020-07-01	2021-01-01	2021-07-01	2022-01-01	2022-07-01	2023-01-01	2023-07-01
1_100	0	0	0	0	0	0	0	0	0	0	0	0	0
101_200	0	0	0	0	0	1	3	0	0	0	0	0	0
201_500	0	4	3	4	3	2	0	1	0	0	0	0	0
501_1000	0	0	0	0	0	0	0	0	5	7	6	3	13
1001_999999	187	174	197	235	269	297	332	335	364	384	364	380	333



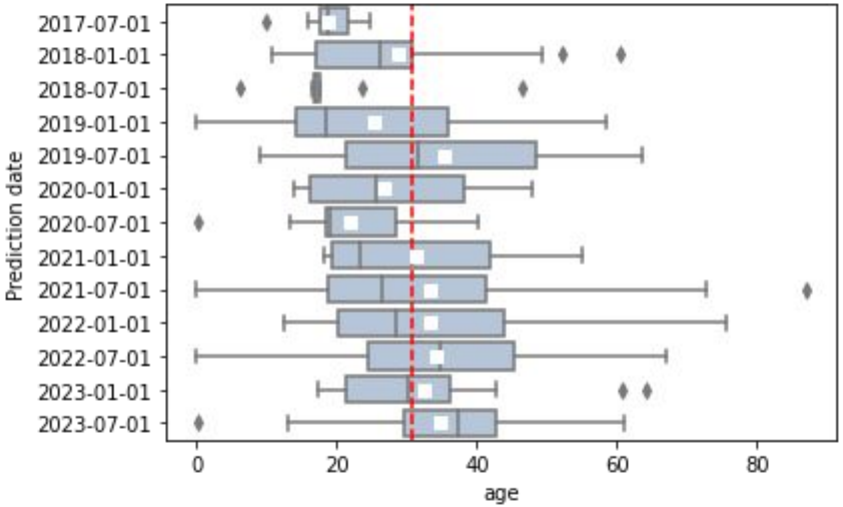
Model comparisons - Demographics ⇒ Age

Flagged by current model missed by Scorecard



Mean: 38 years

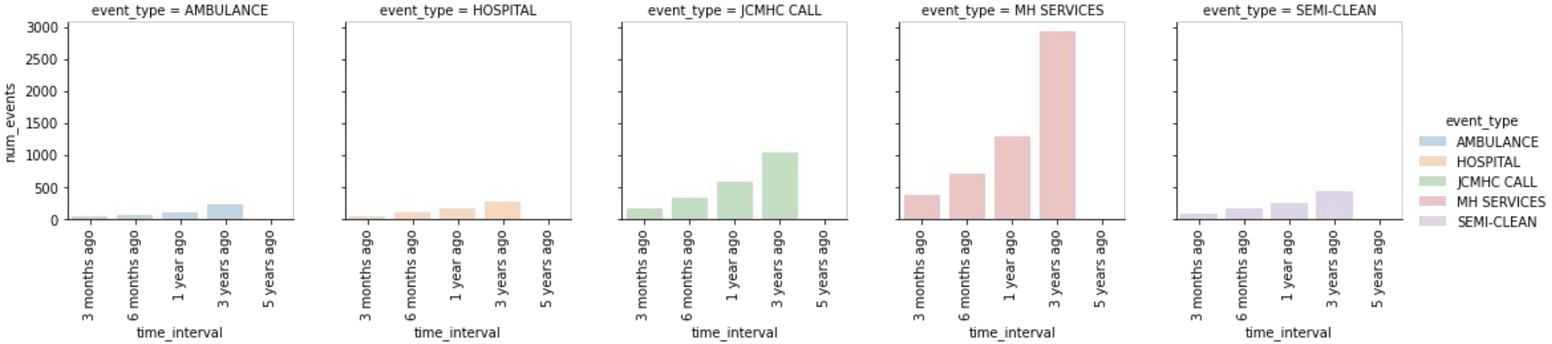
Flagged by Scorecard missed by Current model



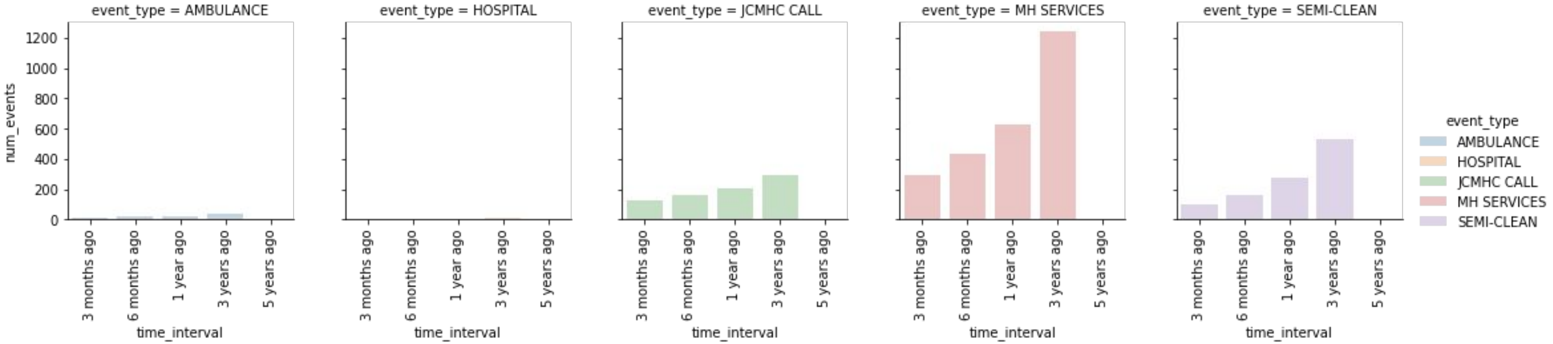
Mean: 31 years

Model comparisons - Type of events

Flagged by current model missed by Scorecard

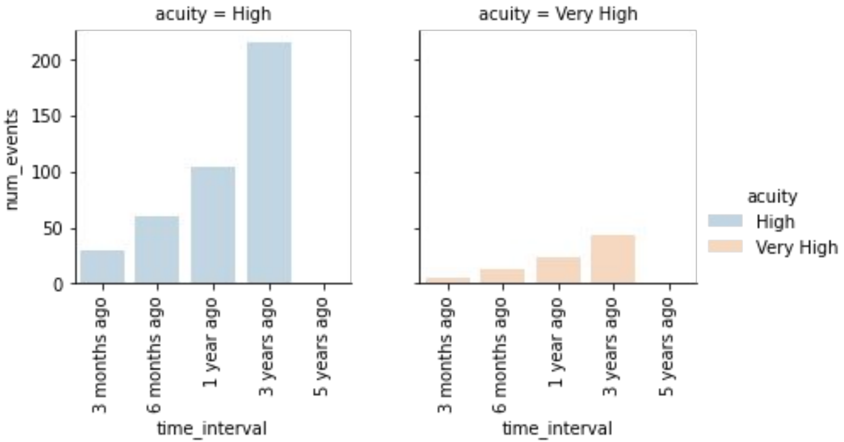


Flagged by Scorecard missed by Current model

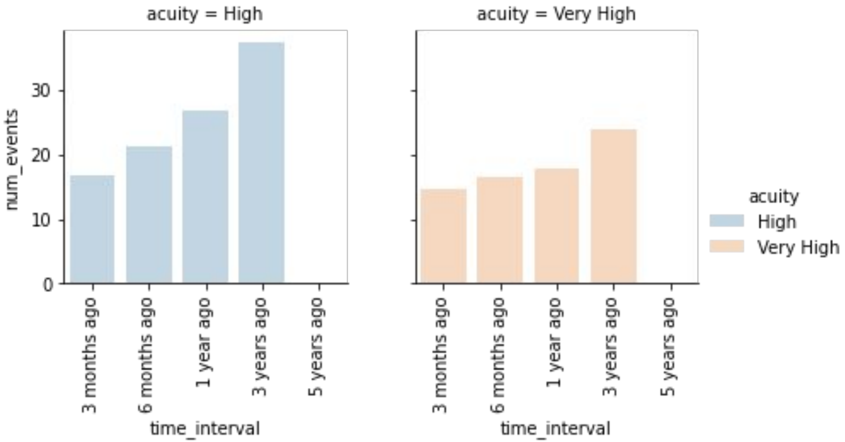


Model comparisons - Acuity of events

Flagged by current model missed by Scorecard



Flagged by Scorecard missed by Current model



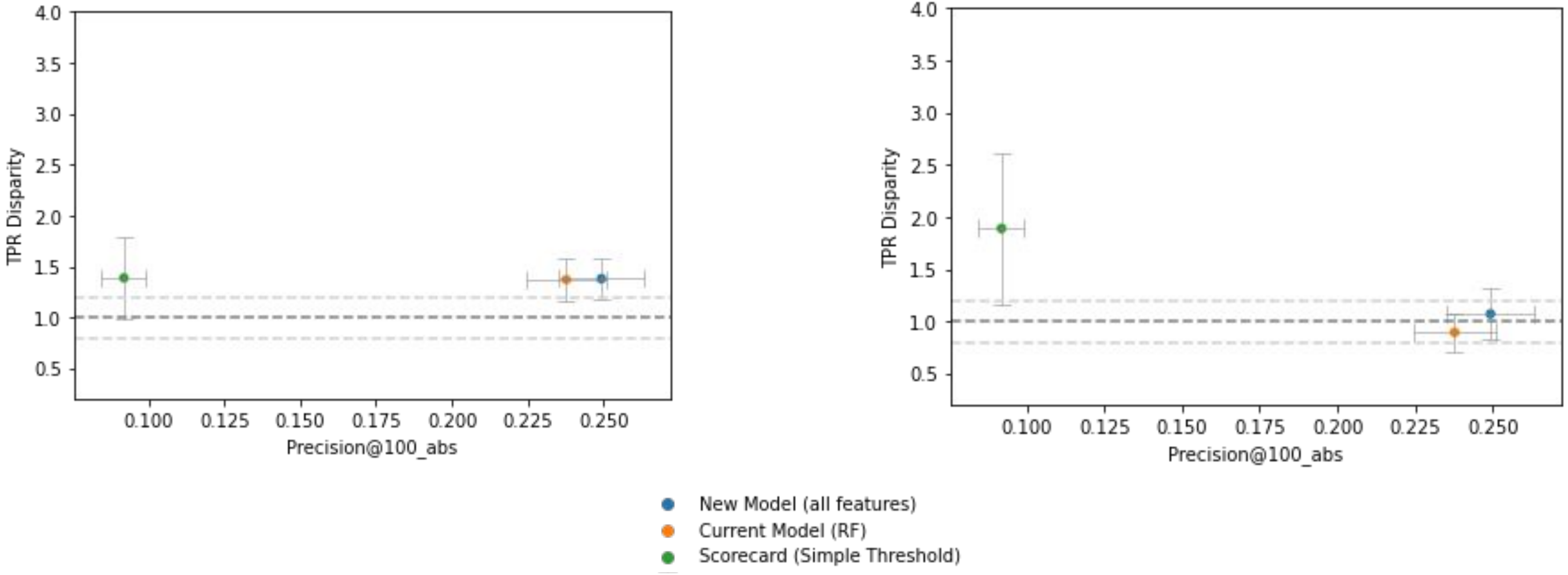
Bias and Fairness: Recall disparity

Gender

Protected ⇒ Female | Reference ⇒ Male

Race

Protected ⇒ Black | Reference ⇒ White



To sum up

- Post-modeling happens once you have a subset of models selected based on performance
- You use the post-modeling analysis to identify and select the model that will be validated with a field trial
- Post-modeling gives information about the entities highlighted by the model of having the highest likelihood of having/experiencing the outcome
- Post-modeling analysis includes several types of analysis mainly to characterize the entities in your top k lists

Responsible AI isn’t just about explainability or bias metrics—it’s about recognizing who our models serve and who they overlook. Our technical decisions have real-world consequences, affecting **individual lives**. As analysts/scientists, we **must be thorough**!