

DATA+AI
SUMMIT 2022

Improving Substance Abuse Treatment Outcomes

Predicting Repeat Admissions with Machine Learning



Kelsey Emmett, MPH
Lead Data Scientist, Kimberly Clark



Jennifer Morizzo
Associate Data Scientist, Maritz

ORGANIZED BY  databricks

The Goal: Personalized Treatment

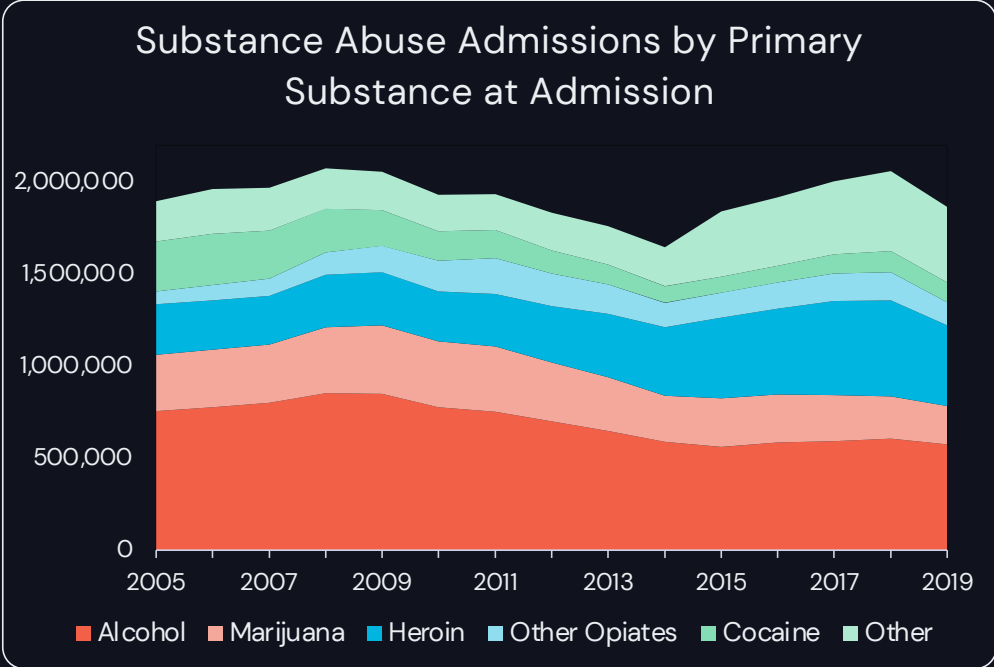
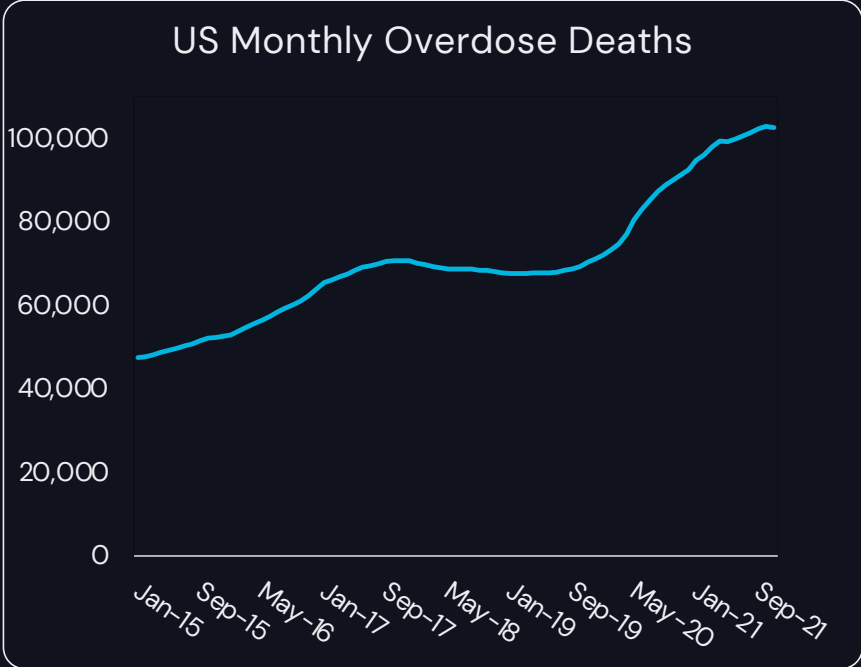
Customize treatment based on predicted outcomes

- Predict with high accuracy patients who will relapse and return to substance abuse treatment
- Use model explainers to find individual-level prediction contributors
- Personalize treatment to maximize success

Background

The Problem: Substance Abuse at a Glance

Overdose deaths and admissions from hard drugs are rising



Contains data from publicly-funded treatment admissions

Treatment Episode Dataset (TEDS)

Model trained with substance use treatment admission data

- Built with publicly available data from the Substance Abuse and Mental Health Services Administration (SAMHSA)
- Includes all state and federally funded substance abuse treatment admissions
 - Model best used on high-risk populations
- Observations at an admission level
 - A limitation because historical behavior cannot be tracked

Target Variable for Prediction:

- Whether an admission is a:
 - First admission
 - Repeat admission

Final Model & Evaluation

Final Model: Random Forest Classifier

Multiple models were tested using a grid search

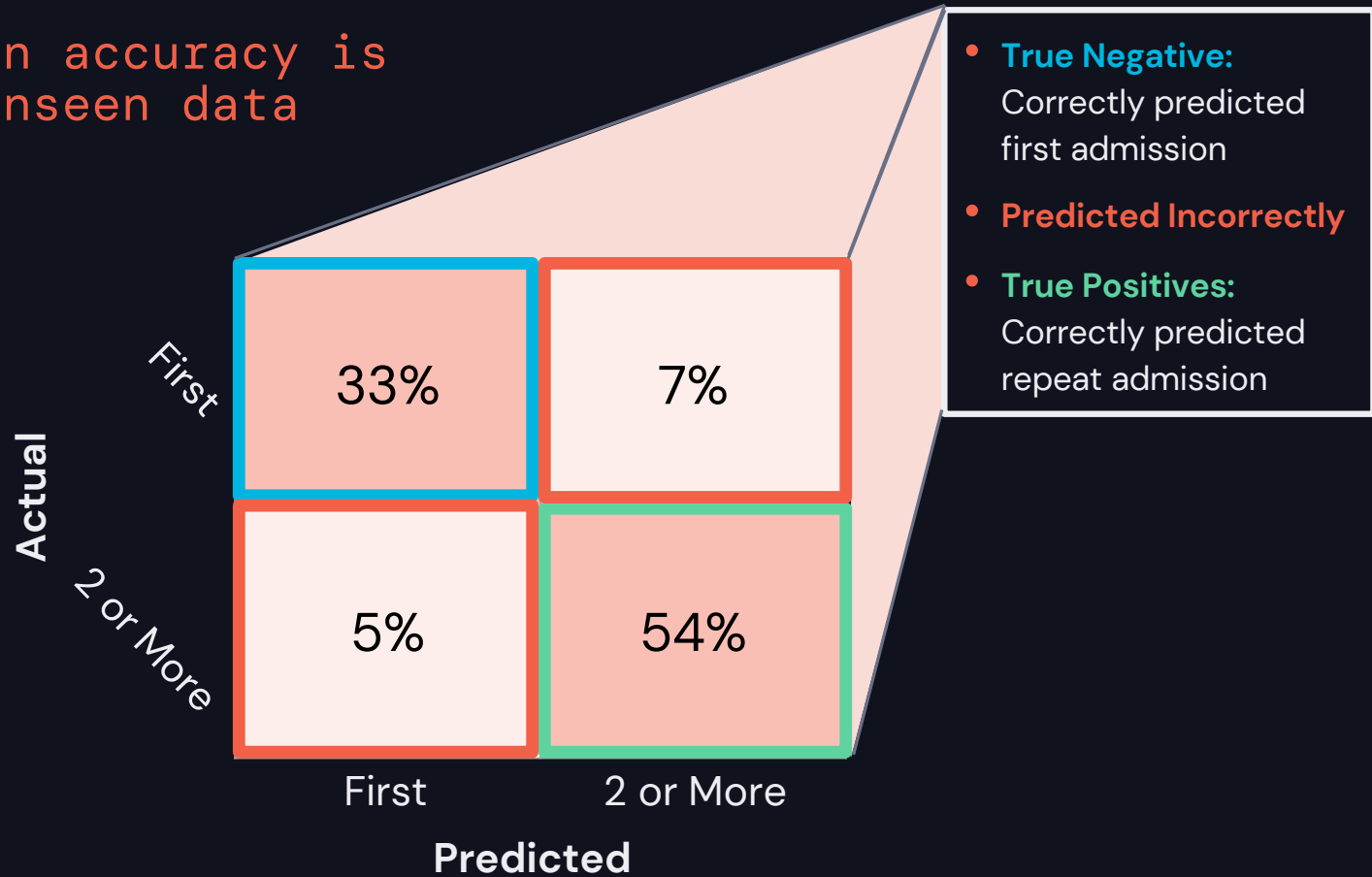
- Tested multiple model types:
 - **Random Forest**
 - Gradient Boosted Trees
 - XGBoost
 - Deep Neural Networks
 - Lasso and Ridge Regression

Final Model Evaluation Metrics

- Accuracy: 0.88
- Precision: 0.89
- Recall: 0.91
- F-Score: 0.90
- AUC: 0.87

Model Evaluation

Prediction accuracy is high on unseen data





Feature Importance

Certain features more effectively predict repeat admissions

- Top two factors associated with higher probability of repeat admission are:
 - Higher education
 - Lower frequency of use
- The highly educated are less likely to complete treatment
- Frequent users are less likely to seek or complete treatment

Feature	Importance
1) Education	8%
2) Frequency of Use	6%
3) Age Group	5%
4) Age of First Use	5%
5) Gender: Female	4%
6) Residing in the Pacific Division	4%

Increasing value = increased probability 

Increasing value = decreased probability 

Technical Deep Dive

Technologies Used

Used a variety of technologies for modeling and dashboards

Databricks

- Used Python on Databricks to do a grid search of 100+ models



Scikit-Learn & TensorFlow

- Trained multiple Scikit-Learn and TensorFlow models



MLFlow

- Tracked models using MLFlow
- Recorded model metrics and logged artifacts for evaluation



Power BI

- Created a Power BI dashboard to:
 - Inform clinical decision making
 - Provide model explainability



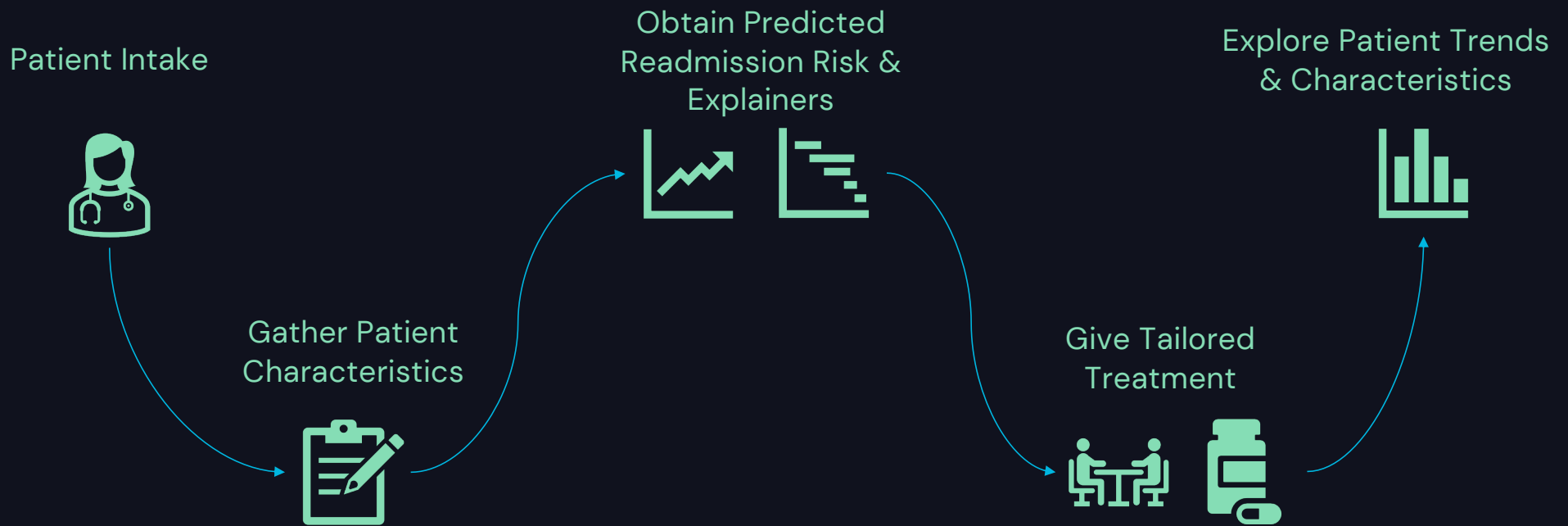
Future Goals of Clinician Solution

Personalize and improve patient treatment

- Predict with high accuracy patients were readmitted to treatment
- Use model explainers to find the most important features contributing to readmission risk
- Personalize treatment to meet patient needs:
 - Vary treatment length to maximize success
 - Focus on addressing factors contributing to high patient predictions

Health Care Provider Journey

An analytics solution for providing personalized treatment

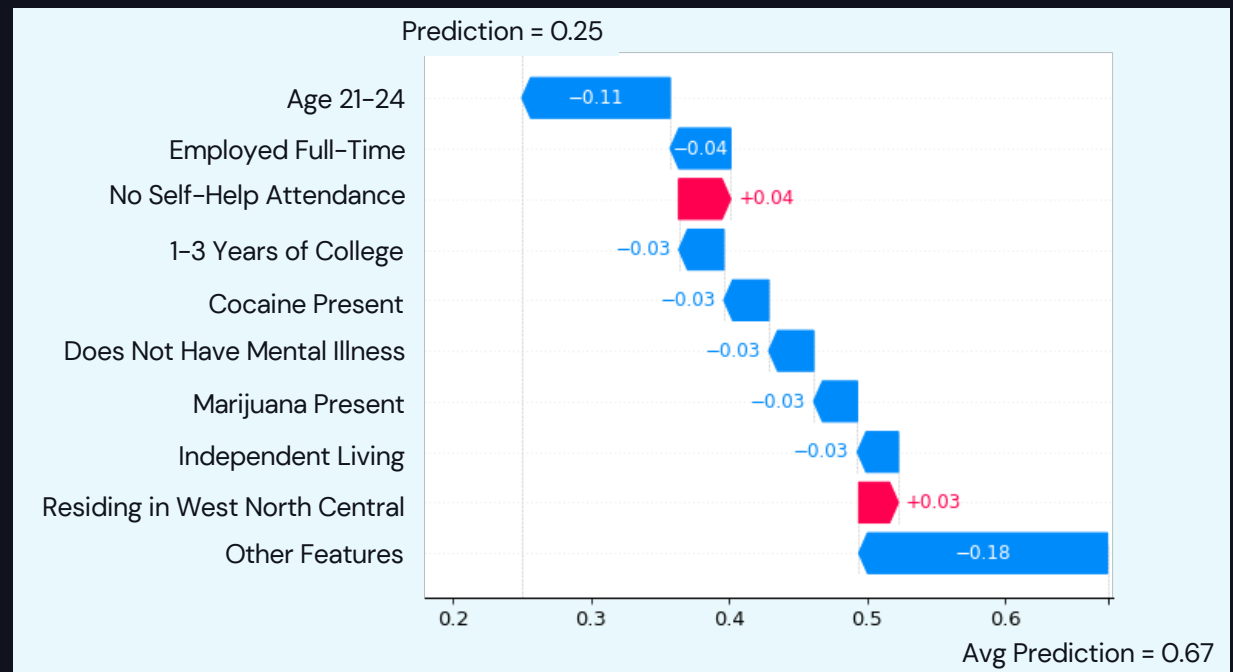


Dashboard Demo

Machine Learning Model Explainers

Explaining individual readmission predictions

- Shapley values quantify the contribution of each feature
- Report feature contributions for each patient
- On right, age group reduced this patient's predicted readmission probability by 0.11



DATA+AI
SUMMIT 2022

Thank you



Kelsey Emmett, MPH
Lead Data Scientist, Kimberly Clark



Jennifer Morizzo
Associate Data Scientist