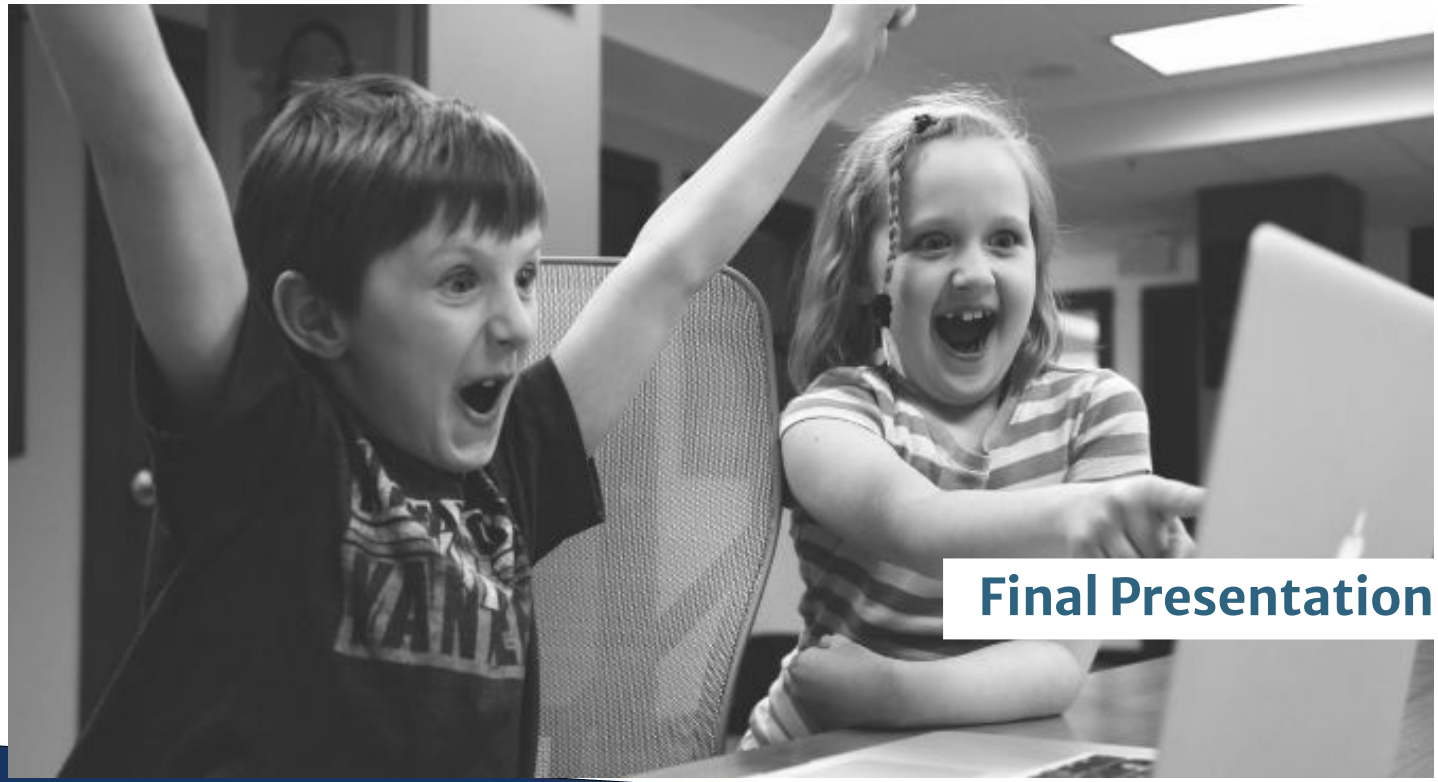
ADHD screening tool



Final Presentation

Joy First
Prashant Dhingra
Sebastian Urbina
Jordan Thomas

W210, Fall 2020

Berkeley
UNIVERSITY OF CALIFORNIA

ADHD – an honest list of fears and confessions

TEDx

Cal State LA

Not Just LIVING

but THRIVING with ADHD

Angela Aguirre

Berkeley
UNIVERSITY OF CALIFORNIA

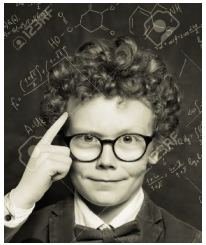
WHY?

support **early diagnosis** of the ADHD

by evaluating the **risk** based on

demographic & **behavioral** factors

Berkeley
UNIVERSITY OF CALIFORNIA



Data sources

National Survey of Children's Health (2017–2020)

109K

Surveys Completed

12K

Diagnosed ADHD

Children (0–17)

Cash Incentives

Random Addresses

Data preparation

Missing Value

Remove values

Feature Selection

One hot encoding

Test/Train Split

Berkeley
UNIVERSITY OF CALIFORNIA

Modeling approach

Predicted value

Actual value	0	1
0	True Positive (TP)	False Positive (FP)
1	False Negative (FN) Key Goal : Reduce error Child has ADHD and Model predict Healthy	True Negative (TN)

Label 0 - Child is healthy, 1 - ADHD

minimize FN & **maximize** recall

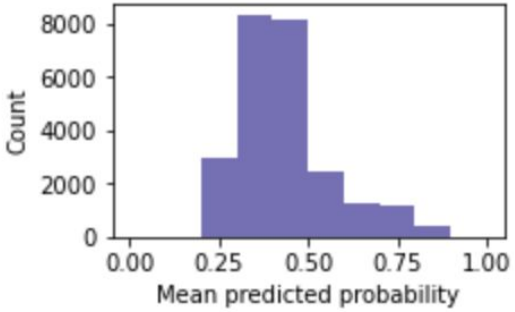
$TP/(TP + FN)$

Selected Model – Random Forest

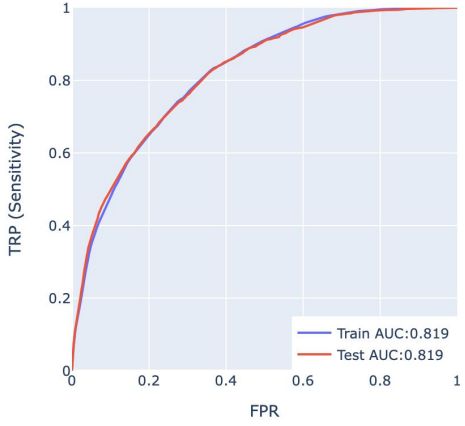
Confusion Matrix

TP 14,920	FP 2818
FN 778	**TN** 1244

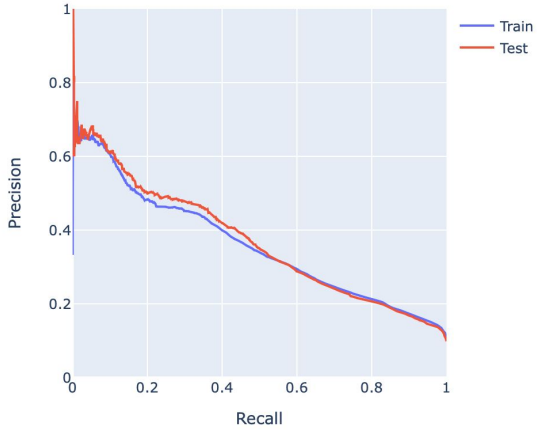
Mean Predicted probability



ROC Curve



PR Curve



Other models

- Gradient Boosted (.9)
- KNN
- Logistic Regression
- Naive Bayes
- Random Forest (.8)
- Random Forest with different layers

Metrics

- Confusion Matrix
- ROC Curve

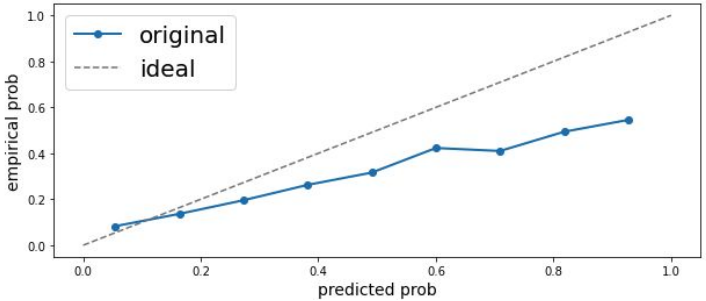
Class Imbalance

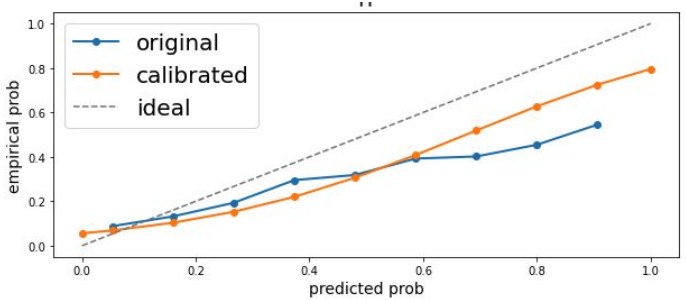
- Mean Predicted Probability
- PR Curve

Model – Training and Inference

RF Classifier → Calibrator

Training



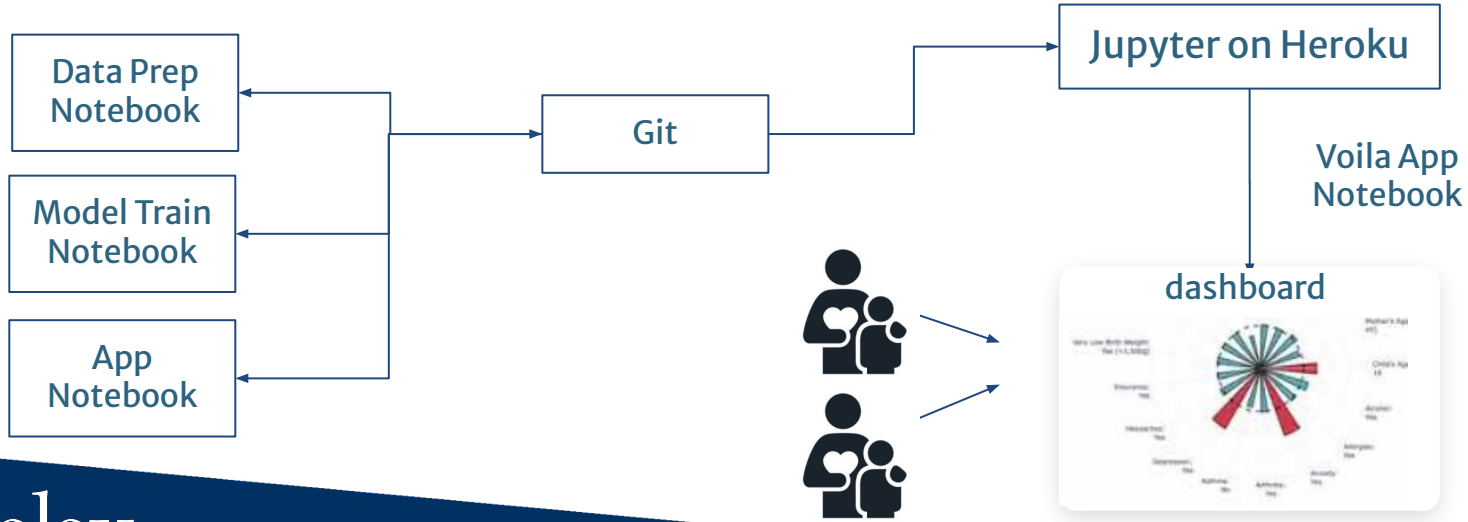


Inference

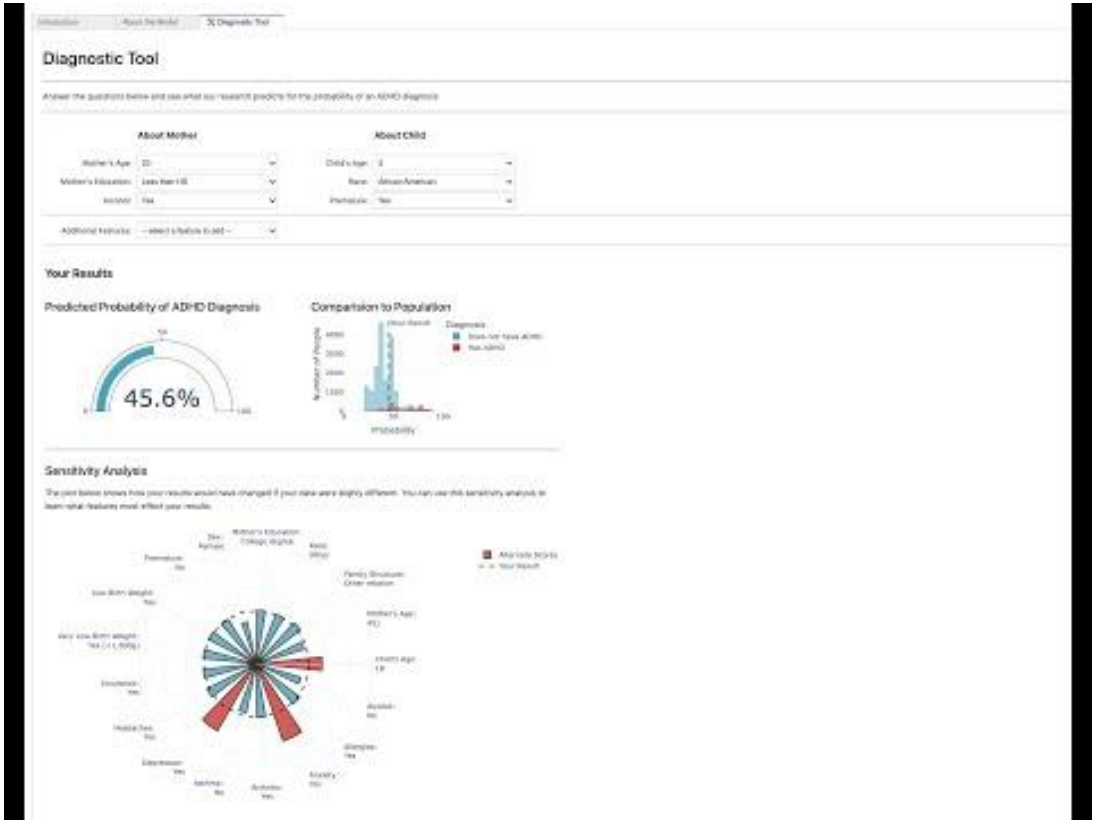
Input → RF Model → .72 → Calibration → .64

How can we share the results?

Publish our notebooks as dashboards for parents



DEMO



Berkeley
UNIVERSITY OF CALIFORNIA

TEDx

Cal State LA

Not Just **LIVING** but **THRIVING** with ADHD

Angela Aguirre

Berkeley
UNIVERSITY OF CALIFORNIA

Thank you

Berkeley
UNIVERSITY OF CALIFORNIA

Appendix

Links

- Application: Link
- Git Repository: Link
- Notebook: Link

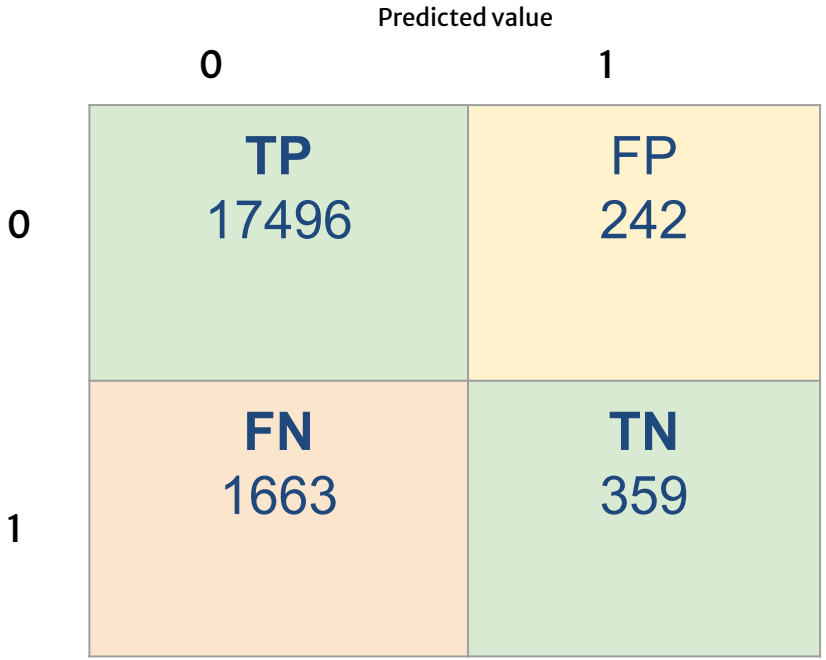
Team Contribution

- **Prashant Dhingra** – *Build multiple modeling architecture - Gradient boosted decision Tree, Random Forest. Built matrix layer to evaluate model e.g. Confusion matrix, ROC, PR, Mean predicted probability. Build calibration curve to reduce over prediction.*
- **Sebastian Urbina** – *Extracted the data set and performed the data wrangling. Developed the pre-processing scripts, to select the features, impute data, clean data and remove data sources. Wrote the documentation for the app. Provide feedback on modelling and UI*
- **Jordan Thomas** – *Helped with initial EDA. Built baseline model pipeline. Built front end application notebooks. Built deployment process for front end application.*
- **Joy First** – *conducted preliminary research (including project option identification, respective datasets, and reference materials), developed a product vision, led the delivery schedule and task assignment, coordinated team meetings and deliverables, conducted subject research and feature selection, and collaborated on data preparation, ML modeling and UX design.*

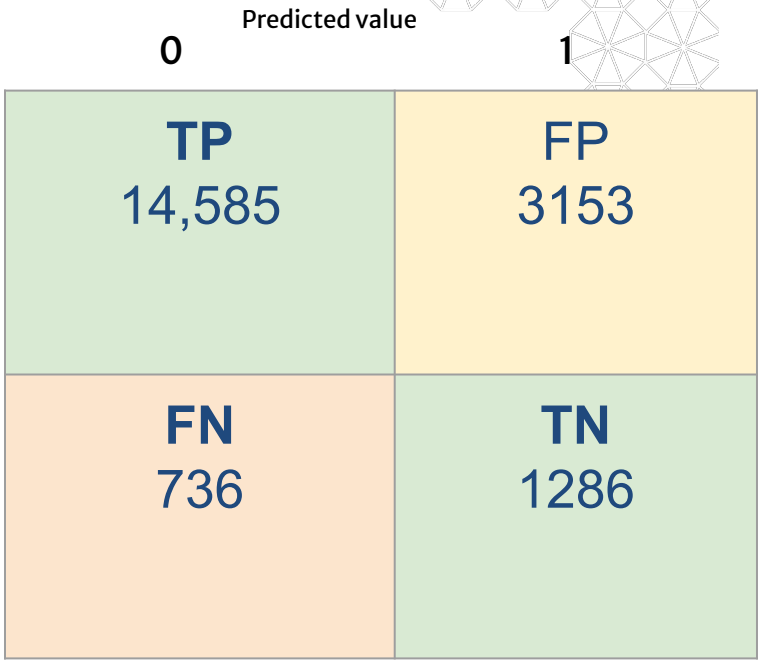
References

- "Economic burden of attention-deficit/hyperactivity disorder among adults in the United States: a societal perspective." PubMed, 22 November 2021, https://pubmed.ncbi.nlm.nih.gov/34806909/. Accessed 28 August 2022
- "Accurate Identification of ADHD among Adults Using Real-Time Activity Data." NCBI, 26 June 2022, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9312518/. Accessed 28 August 2022.
- "Predicting Children with ADHD Using Behavioral Activity: A Machine Learning Analysis." MDPI, https://www.mdpi.com/2076-3417/12/5/2737/htm. Accessed 28 August 2022.

Gradient Boosted Decision Tree



Random Forest



GB has .90 and RF has .82 accuracy.
But RF is better for FN