ADHD screening tool
ADHD – an honest list of fears and confessions

TEDx Cal State LA
Not Just LIVING but THRIVING with ADHD
Angela Aguirre
support **early diagnosis** of the ADHD by evaluating the risk based on demographic & behavioral factors
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
**Modeling approach**

<table>
<thead>
<tr>
<th>Predicted value</th>
<th>True Positive (TP)</th>
<th>False Positive (FP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>True Positive (TP)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>1</td>
<td>False Negative (FN)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

Key Goal: Reduce error

Child has ADHD and Model predict Healthy

Label 0 – Child is healthy, 1 – ADHD

- **minimize** FN
- **maximize** recall

\[ \text{minimize } FN \quad \text{and} \quad \text{maximize } \text{recall} \]

\[ \frac{TP}{(TP + FN)} \]
Selected Model – Random Forest

Confusion Matrix

- TP: 14,920
- FP: 2818
- FN: 778
- TN: 1244

Mean Predicted probability

ROA 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

- Data Prep Notebook
- Model Train Notebook
- App Notebook

Git

Jupyter on Heroku

Voila App Notebook

dashboard

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

*Angela Aguirre*
Thank you
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

<table>
<thead>
<tr>
<th></th>
<th>Predicted value</th>
<th></th>
<th>Predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TP</td>
<td>17496</td>
<td>FP</td>
<td>242</td>
</tr>
<tr>
<td>FN</td>
<td>1663</td>
<td>TN</td>
<td>359</td>
</tr>
</tbody>
</table>

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