ADHD Risk Prediction Using Multimodal Neuroimaging and Behavioral Data

[4pt]

Author: Aman Shah
GitHub: @ASdata1

A machine learning pipeline that predicts **ADHD risk** using functional neuroimaging connectomes, behavioral questionnaires, and demographic data. The project demonstrates full end-to-end ML development from data preprocessing and feature engineering to model evaluation and deployment.

Overview

- **Dataset:** 872 patients, 19,900+ fMRI connectivity features, and behavioral scales (SDQ, APQ, EHQ)
- **Objective:** Classify ADHD vs. non-ADHD using multimodal inputs
- **Model:** Regularised Logistic Regression with PCA and KNN imputation
- **Performance:** ROC-AUC = 0.802, Precision (ADHD) = 85.1%, Recall (ADHD) = 82.5%, stable across 10 seeds ($\pm$0.004 std)

Methodology

Preprocessing: Standard scaling before KNN imputation to avoid distance bias, one-hot encoding for categorical variables, PCA reducing 19,900 features to 10 principal components.

Model Development: Logistic Regression with L1/L2 regularisation, GridSearchCV hyperparameter tuning, class balancing (`class_weight='balanced'`), and custom threshold optimisation for best F1-Macro (0.45).

Validation: Stratified 60/20/20 split with final evaluation on an untouched test set.

Key Insights

- Combining neuroimaging and behavioral data improves ADHD screening accuracy.
- High recall ensures minimal missed cases—suitable for pre-clinical triage.
- Feature scaling and PCA sequencing significantly influenced performance.

Tech Stack

Python, Scikit-learn, Pandas, NumPy, Matplotlib, Streamlit, Joblib

Repository Structure

```
ADHD_Prediction/
 notebooks/              # EDA, feature engineering, modelling
 models/                 # Saved artifacts (.joblib, .json)
 app.py                  # Streamlit demo
 requirements.txt
 README.md
```

Run Locally

```
git clone https://github.com/ASdata1/ADHD_SEX_Prediction.git
cd ADHD_SEX_Prediction
pip install -r requirements.txt
streamlit run app.py
```

Clinical Relevance

The model serves as a screening aid to prioritize children for ADHD assessment, supporting early diagnosis and reducing unnecessary clinical evaluations. With high recall and precision, it demonstrates potential for use in pre-clinical triage and healthcare resource optimisation.

GitHub: https://github.com/ASdata1/ADHD_SEX_Prediction