Brain activity pattern differences in ADHD

Stage 1. Data Exploration

Diego Armando Salinas Lugo ds24353 2401168

The dataset has been explored and preprocessed in Stage 1, by loading and inspecting datasets, handling missing values, merging datasets, performing exploratory data analysis (EDA), identifying correlations and distributions of key features and saving two datasets, one for validation and another one for training with which the exploration has been made.

For Stage 2, the following steps are considered:

- Apply feature scaling for numerical attributes with Min-Max Normalization or Z-score Standardization, since certain models are sensitive to the scale of numerical features.
- Train models.
- Compare models' performance and select the most suitable approach.
- Interpret results to ensure model reliability.

Machine Learning considered models:

- **Logistic Regression**: Since it is a linear model which can estimate the probability of ADHD presence based on the features. It can also provide easily interpretable coefficients, allowing to understand which features contribute the most to ADHD classification.
- **k-Nearest Neighbors (k-NN)**: ADHD classification could benefit from k-NN due to similar behavioral or demographic patterns among participants.
- **Random Forest:** ADHD diagnosis is influenced by different behavioral, demographic, and cognitive features. Random Forest's capability to handle mixed data types makes it robust for this dataset. Besides, it automatically captures feature importance, which can help to get insights into the features influence.
- **Support Vector Machine (SVM):** ADHD diagnosis may not be linearly separable, meaning a model that finds non-linear decision boundaries could be very helpful. SVMs can identify complex patterns, improving classification accuracy.
- **Neural Networks**: As ADHD diagnosis is a multi-faceted problem with diverse variables interacting in non-trivial ways. NN can automatically detect these interactions, making them ideal for capturing complex relationships.

Evaluation Metrics considered

- Accuracy as a general performance measure.
- Precision & Recall to address potential class imbalance issues.
- AUC-ROC To evaluate model discrimination ability.