Predicting Problematic Internet Use in Children

Team #103: Fidel Escalona Badia, David Gilchrist, Ngoc Tu Pham, Daniel Rodriguez, Casey Hatfield

Summary

Problematic internet use (PIU) is contributing to a mental health crisis in adolescents. Traditional PIU assessment depends on clinical evaluations or self-reports, limiting accessibility and early intervention. We created a machine-learning model based on wearable sensor data to predict PIU based on activity levels. We found modest predictive ability in this approach.

What Is PIU?

Problematic Internet Use (PIU) refers to excessive or uncontrolled internet use that negatively impacts an individual's daily life, mental health, and social functioning. Commonly affecting adolescents, PIU is associated with issues such as anxiety, depression, and poor academic performance.

Data Preprocessing

We analyzed both tabular data with demographic, physical, and behavioral attributes, and time-series data from actigraphy collected over 30 days. Tabular data was cleaned, with feature engineering applied to create combined metrics (e.g., BMI_Age, Internet_Hours_Age) for better predictive power. Time-series data was processed by removing incomplete days, aggregating features (e.g., mean and standard deviation) over hourly intervals per week, and reducing dimensionality with Principal Component Analysis (PCA).

Modeling Approach

The models used include LightGBM, XGBoost, and Random Forest and an ensemble method. The model training pipeline was implemented in a Jupyter Notebook. Bayesian optimization was used to fine-tune model hyperparameters, aiming to maximize performance on the quadratic weighted kappa metric.
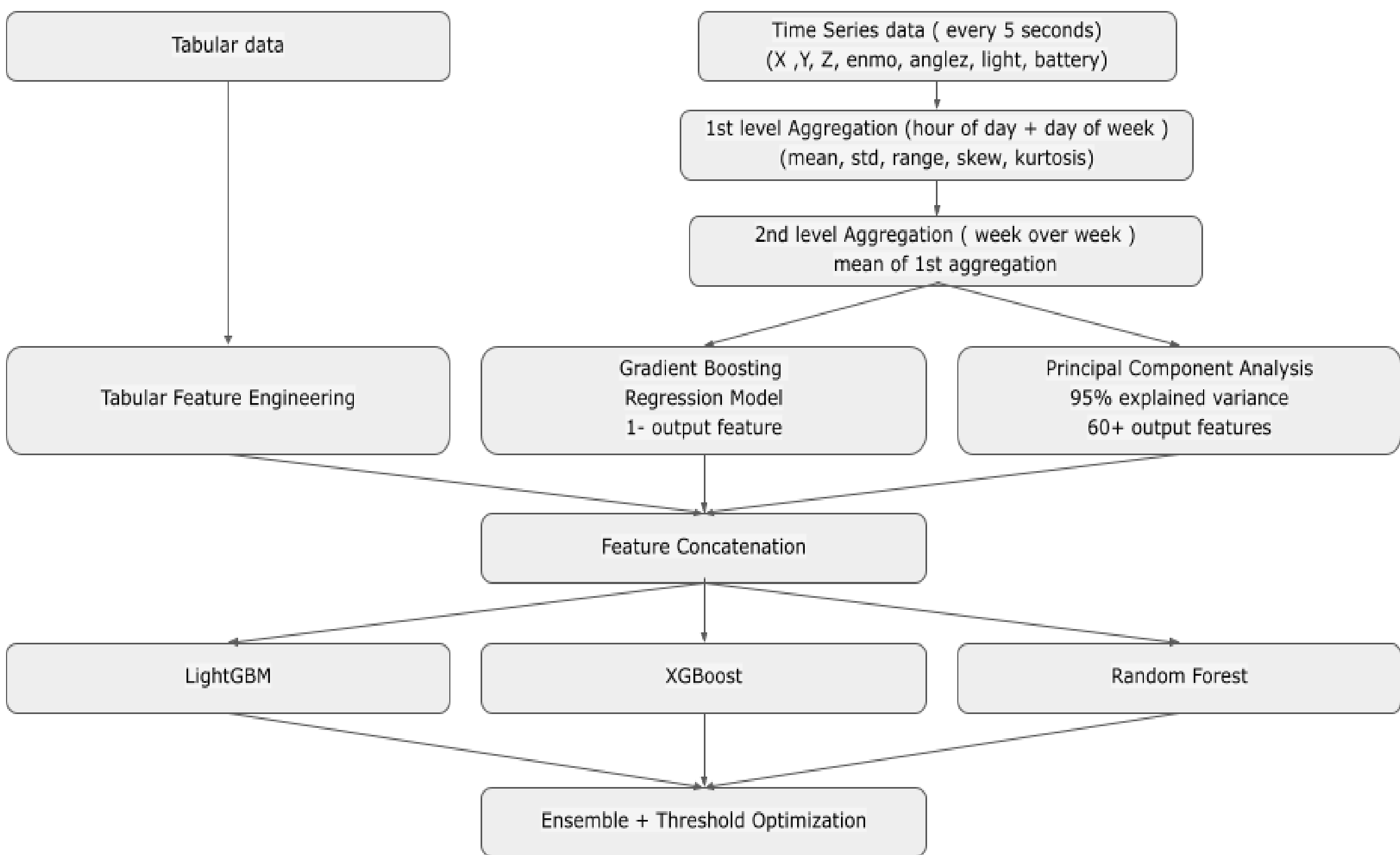
Exploratory Data Analysis
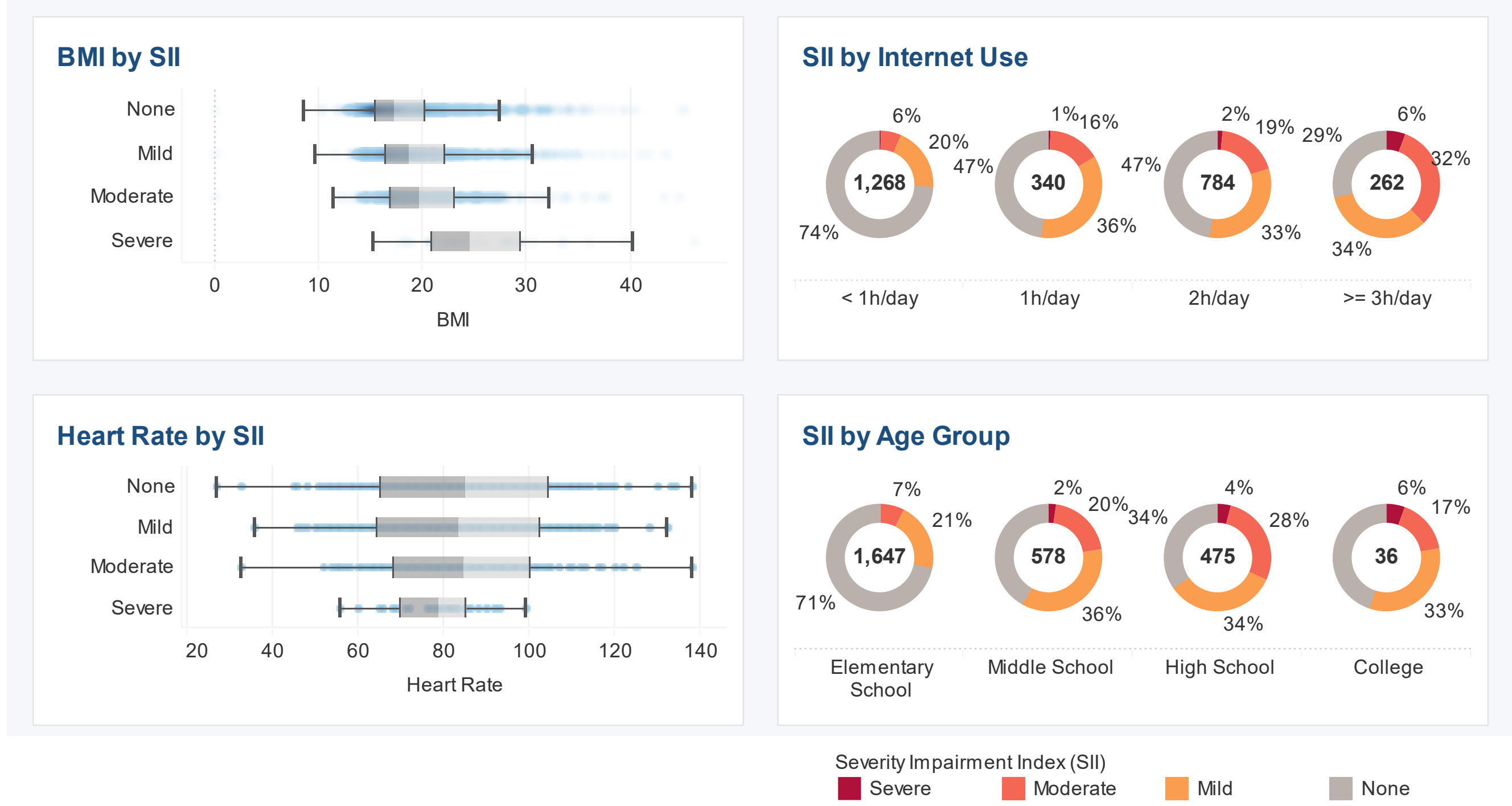
Model Feature Selection
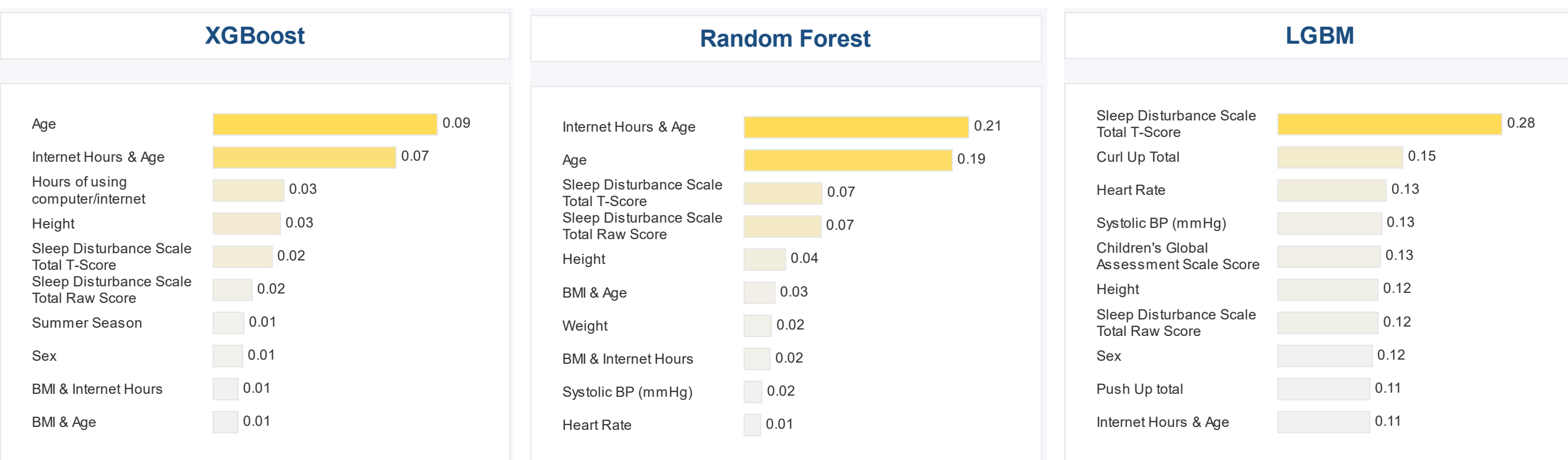
Performance of Ensemble Model

Confusion Matrix

SII	None	Mild	Moderate	Severe
None	1,139	373	82	
Mild	283	305	142	
Moderate	78	161	139	
Severe		5	29	0

SII Pred

Conclusion/Discussion

This study demonstrates that activity levels and demographic data can be used to predict PIU, minimizing reliance on self-reported or expert evaluations. Despite challenges with data quality the results indicate that key physical metrics like sleep disturbance are significant predictors. Future work could focus on refining data collection, preprocessing and balancing techniques to further improve prediction accuracy, showing promise for accessible PIU screening tools in real-world settings.