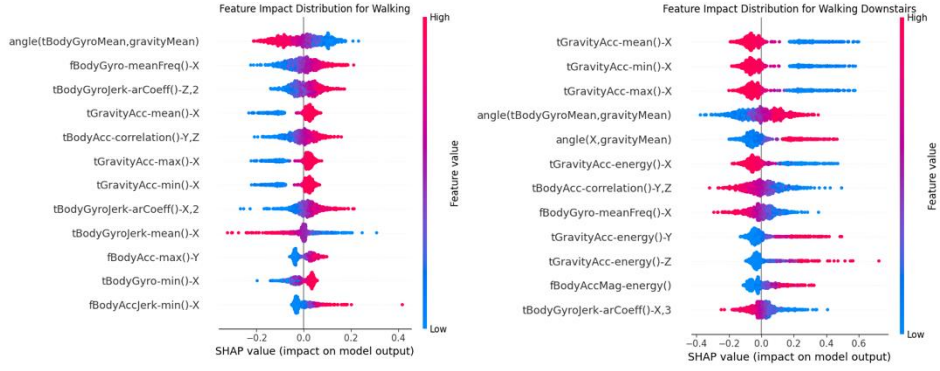
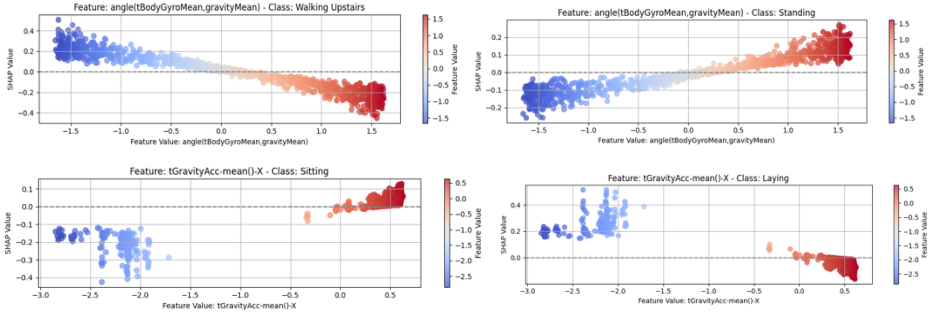
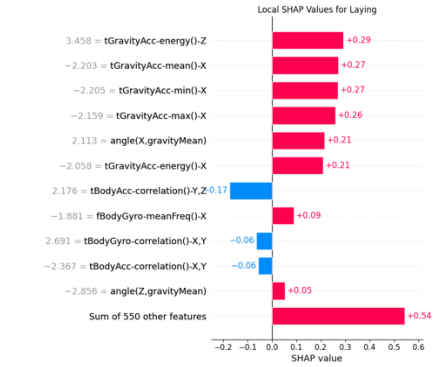
Explaining ML Predictions with SHAP

Avik Basu







Human Activity Recognition (1D-CNN)

Task: Activity classification (6 classes)

Data: 563 features from smartphone sensors

Sample Insights:

- Walking is identified by a smaller angle between the body-gyroscope mean and mean gravity vector
- Reversed behavior is observed for Walking upstairs
- Gravity X-component extremes separate laying and sitting.

Fair, Consistent Attribution

SHAP uses Shapley values from game theory to assign each feature a "fair share" of the model's prediction.

Global + Local Views

Quickly see which features drive overall model behavior (global) and why a specific prediction was made (local) using the *shap* Python library.

Model-Specific Speedups

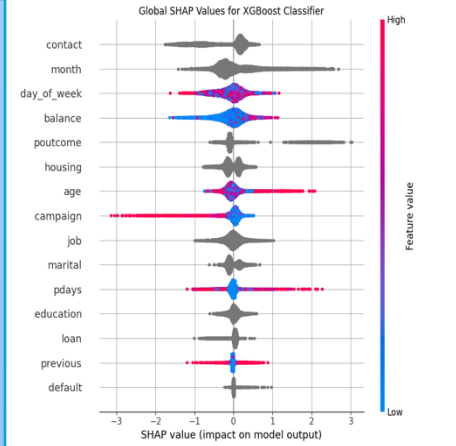
Tree SHAP and Deep SHAP algorithms make otherwise intractable Shapley calculations practical for large tree ensembles and deep neural networks.

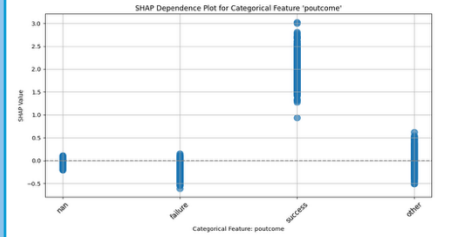
Actionable Insights

Identify high-impact features and uncover when and why a model might behave unexpectedly

Increases Transparency

By visualizing feature contributions, SHAP helps build user and stakeholder confidence and assists in ethical AI review.





Bank Marketing Dataset (XGBoost)

Task: Predict term deposit subscription

Data: 45k rows, 15 features

Sample Insights:

- High account balance drives subscriptions
- Diminishing returns on repeated calls
- Successful previous outreach helps with new deposits