Stacked Ensemble Model (95.71% Accuracy) steps:

1. Data Preparation

Merged Travel and Survey data on ID

Cleaned missing values and encoded categorical features

Scaled numerical data using StandardScaler

2. Model Selection (Base Learners)

The base layer consists of:

CatBoostClassifier

XGBoostClassifier

LightGBMClassifier

RandomForestClassifier

Each was individually tuned using Optuna for hyperparameter optimization.

3. Meta-Learner (Final Estimator)

CatBoostClassifier was used again here, but trained on the predictions of the base models (stacking logic).

This learner combines the strengths of all base models and generalizes well.

4. Evaluation

Cross-validation used to prevent overfitting

SHAP used for model explainability

ROC, confusion matrix, and feature importance plotted