Summary of Data Processing Steps:

1. Data Loading: Loaded the dataset containing 58 features along with the target variable.

2. Outlier Removal: Identified and removed outliers using the Interquartile Range (IQR) method to ensure data quality.

3. Train-Test Split: Split the dataset into training and testing sets for model evaluation.

4. Feature Scaling: Applied feature scaling (StandardScaler) to standardize the range of feature values, ensuring uniformity in model training.

5. Model Training and Evaluation
   - Utilized XGBoost for classification tasks due to its effectiveness with structured/tabular data and ability to handle large datasets efficiently.
   - Employed various feature selection methods, including Variance Threshold, Correlation Coefficient, ANOVA F-test, Forward Selection, Backward Elimination, Lasso (L1 Regularization), Ridge (L2 Regularization), Tree-based feature importance (Random Forest), and Hierarchical Clustering for Feature Selection.
   - Trained models using selected features and evaluated their performance using metrics such as accuracy, precision, recall, and F1 score.
   - Utilized a combination of filter, wrapper, and embedded feature selection methods to identify the most informative features for modeling.

6. Model Optimization and Selection
   - Optimized model parameters using techniques such as grid search and cross-validation to improve performance.
   - Selected the best-performing model based on evaluation metrics and feature selection method.
best model vas lassoL1
7. Test Dataset Prediction:
   - Loaded the test dataset and preprocessed it by applying the same feature scaling and outlier removal techniques used on the training data.
   - Utilized the best-performing model to predict the target variable on the test dataset.
   - Generated predictions and saved them along with corresponding test IDs into a submission file for evaluation.