**Key Improvements and Fixes:**

1. **Structural Organization**:
   * Modularized code into logical sections
   * Used proper function encapsulation
   * Implemented consistent naming conventions
2. **Data Handling Fixes**:
   * Fixed data leakage in normalization using caret::preProcess
   * Standardized data generation pipelines
   * Added proper validation dataset handling
3. **Model Training Enhancements**:
   * Implemented parallel processing
   * Added cross-validation with trainControl
   * Standardized model training interface
   * Added proper metric tracking (ROC-based)
4. **Evaluation Improvements**:
   * Unified evaluation framework
   * Added comprehensive performance metrics
   * Implemented proper visualization functions
   * Fixed ROC curve generation
5. **Code Quality**:
   * Removed redundant code blocks
   * Implemented functional programming patterns
   * Added proper error handling
   * Improved documentation and readability
6. **Critical Bug Fixes**:
   * Corrected validation data normalization
   * Fixed class imbalance in original dataset
   * Addressed improper feature scaling
   * Resolved random seed management issues
7. **Performance Optimizations**:
   * Implemented parallel processing
   * Optimized data generation
   * Reduced memory footprint
   * Improved computational efficiency

**Recommendations for Further Improvement:**

1. Implement hyperparameter tuning grids
2. Add feature selection mechanisms
3. Incorporate class imbalance handling
4. Add automated reporting
5. Implement unit testing
6. Add model serialization/deserialization
7. Implement Shapley value interpretation
8. Add Docker containerization
9. Implement MLflow experiment tracking

This version provides a more robust framework for comparing normality assessment methods while maintaining scientific rigor and computational efficiency.

New chat