1. Ordinal Encoding for Age Groups

- **Current Approach:** We used ordinal encoding for age groups, assigning integer labels (e.g., 1, 2, 3) to represent the groups. However, this approach assumes a uniform difference between groups, which might not hold true since the ranges of the groups are unequal.
- **Suggested Improvement:** Instead of ordinal encoding, we could represent each group by its **average age** (e.g., replacing the group "20–30" with 25). This would better capture the numerical relationship between age groups and align them with the model's assumptions for continuous features.

2. Feature Selection with Spearman's Correlation

- **Current Issue:** In our feature selection process, we excluded one of the features in pairs of highly correlated features, selecting the one with higher variance. However, since we applied **StandardScaler**, all features have been normalized to have the same variance (1), making this criterion ineffective.
- **Suggested Improvement:** We should reconsider the selection criteria for correlated features. For example, retain the feature with higher **importance to the target variable** (e.g., based on mutual information or feature importance from a preliminary model).

3. Hyperparameter Space Expansion to Reduce Overfitting

- **Current Approach:** Our model may be overfitting due to insufficient exploration of the hyperparameter space during optimization. A limited hyperparameter space restricts the model's ability to find an optimal balance between underfitting and overfitting.
- **Suggested Improvement:** Increase the hyperparameter space by exploring a wider range of values for key parameters (e.g., learning rate, number of estimators, tree depth).