Stratified KFold Cross-validation

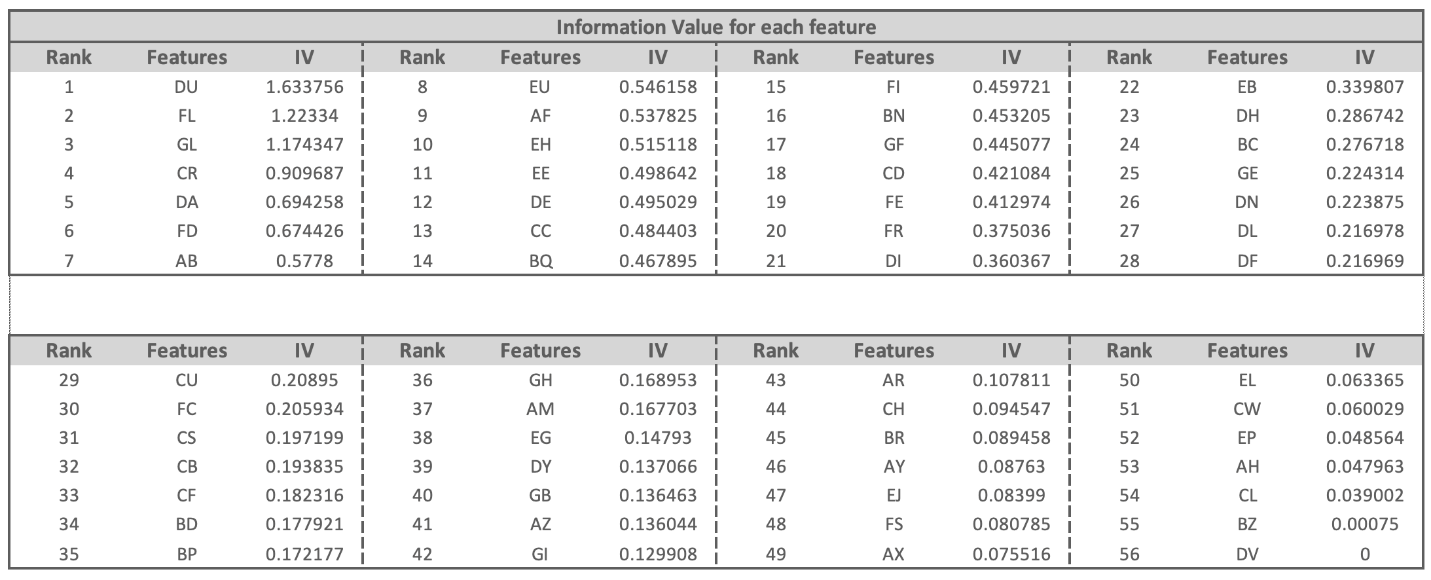
Stratified KFold Cross-validation is a special type of cross-validation technique used to evaluate the performance of a model in an imbalanced dataset. It uses stratified sampling technique to select the samples in the same proportion as the KFold image appear in the population

K = 10 is chosen as the number of folds in each iteration to balance the bias and variance.

Information Value

“Information Value (IV) is a metric to measure the predictive power of a feature about the target value. The IV is calculated by comparing the distribution of the feature values for the good and bad outcomes. The IV value for a feature can range from 0 to infinity, with higher values indicating stronger predictive power. The formula to calculate IV is:

The IV of these 56 features in the dataset is presented below in descending order.



The IV is used in conjunction with the correlation values and variance inflation factors (discussed in the next part) to build a systematic feature selection method. “