



Figure 2. Default ML Pipeline for AutoML. The default ML pipeline can be represented as a flowchart that starts by splitting the input dataset into training and independent test sets, provided the user has activated the test set option. Otherwise, the entire dataset is used for training. In the subsequent step, the training dataset is divided into n bins of equal size through stratified sampling. From these bins, k-1 are designated as training sets while the remainder becomes the test set. In the pre-processing step, low variance features are removed first, followed by data scaling and resampling. Subsequently, the SelectPercentile univariate feature selection method is applied to select important features, and five ML classification algorithms are trained. Model performance is assessed on the test set using three different methods, and multiple performance metrics are computed. This entire process is repeated for each unique bin in the k-fold CV method. The pipeline outputs a zip file comprising the log .txt and the results.pkl files. The user can examine the results by visualizing the contents of the pickle file using Machine Learning Made Easy (MLme).