Highlight Note

Configuration and Featurization

You can use the following configuration settings to control the automated machine learning experiment:

- The primary metric for which you want to optimize the model – for example, **Accuracy**. Different task types support different metrics.
- Whether or not to generate feature importance explanations for the best model.
- Blocking algorithms that you do not want the experiment to try.
- Exit criterion to stop the experiment after a maximum amount of time or when a specific metric threshold is achieved.
- The validation technique used to split training and test data to evaluate model performance.
- The number of concurrent training runs.

Automated machine learning also supports data pre-processing, or featurization. It always tries various data normalization/scaling techniques and applies data guardrails to mitigate unbalanced data (for example, datasets where there are a significantly larger number of observations with one label value than another). You can optionally apply additional featurization techniques to individual columns, including dropping high-cardinality features (which tend not to be predictive), imputing missing values, encoding categorical features, and deriving features (for example, by splitting dates into day, month, and year features).