2015 Branco A Survey of Predictive Modelling Under Imbalanced Distributions

Preprocessing of data to account for imbalance

* Resampling
* Active learning – selecting best samples to learn and omitting poor ones
* Weighting data space – modify training dataset using information concering misclassification costs

Resampling

* Oversampling, can result in overfitting
* Undersampling, can lose data points