

# Summary of Text Categorization Methods

- Many methods are available, but no clear winner
  - All require effective feature representation (need domain knowledge)
  - It is useful to compare/combine multiple methods for a particular problem
- Most techniques rely on supervised machine learning and thus can be applied to **any** text categorization problem!
  - Humans annotate training data and design features
  - Computer optimizes the combination of features
  - Good performance requires 1) effective features and 2) plenty of training data
  - Performance is generally (much) more affected by the effectiveness of features than by the choice of a specific classifier