

Candidate study
Title: Defect Prediction
Abstract: Apply SVM to defect prediction.

Tokenization
Concatenating title and abstract,
stemming and stop words removal

Bag of words
["defect", "predict", "appli", "svm", "defect", "predict"]

Featurization
Term Frequency

Term Frequency
{ "defect": 2, "predict": 2, "appli": 1, "svm": 1 }

Dimensionality Reduction
Feature Hashing

Feature Vector
[3, 2, 1]

Normalization
L2 normalization

Normalized Feature Vector
[0.80, 0.53, 0.27]

PREPROCESSING