**PREPROCESSING**

**Normalization**

L2 normalization

**Tokenization**

Concatenating title and abstract,

stemming and stop words removal

**Feature Vector**

[3, 2, 1]

**Normalized Feature Vector**

[0.80, 0.53, 0.27]

**Dimensionality Reduction**

Feature Hashing

**Featurization**

Term Frequency

**Term Frequency**

{“defect”: 2, “predict”: 2, “appli”: 1, “svm”: 1}

**Bag of words**

[“defect”, “predict”, “appli”, “svm”, “defect”, “predict”]

**Candidate study**

**Title:** Defect Prediction

**Abstract:** Apply SVM to defect prediction.