**Normalization**

L2:

**Normalized Feature Vector**

[0.67, 0.67, 0.33]

**Dimensionality Reduction**

Feature Selection by tf-idf

**Feature Vector**

[2, 2, 1]

**Featurization**

Term Frequency

**Tokenization**

Concatenating title and abstract,

stop words removal

**Bag of words**

[“defect”, “prediction”, “apply”, “svm”, “defect”, “prediction”]

**Term Frequency**

{“defect”: 2, “prediction”: 2, “apply”: 1, “svm”: 1}

**Candidate study**

**Title:** Defect Prediction

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