**Exercise 1. Make an ARFF file from the labeled mini-documents in Table 1 and run StringToWordVector with default options on this data. How many attributes are generated? Now change the value of the option minTermFreq to 2. What attributes are generated now?**

* Train documents:

A screenshot of a computer

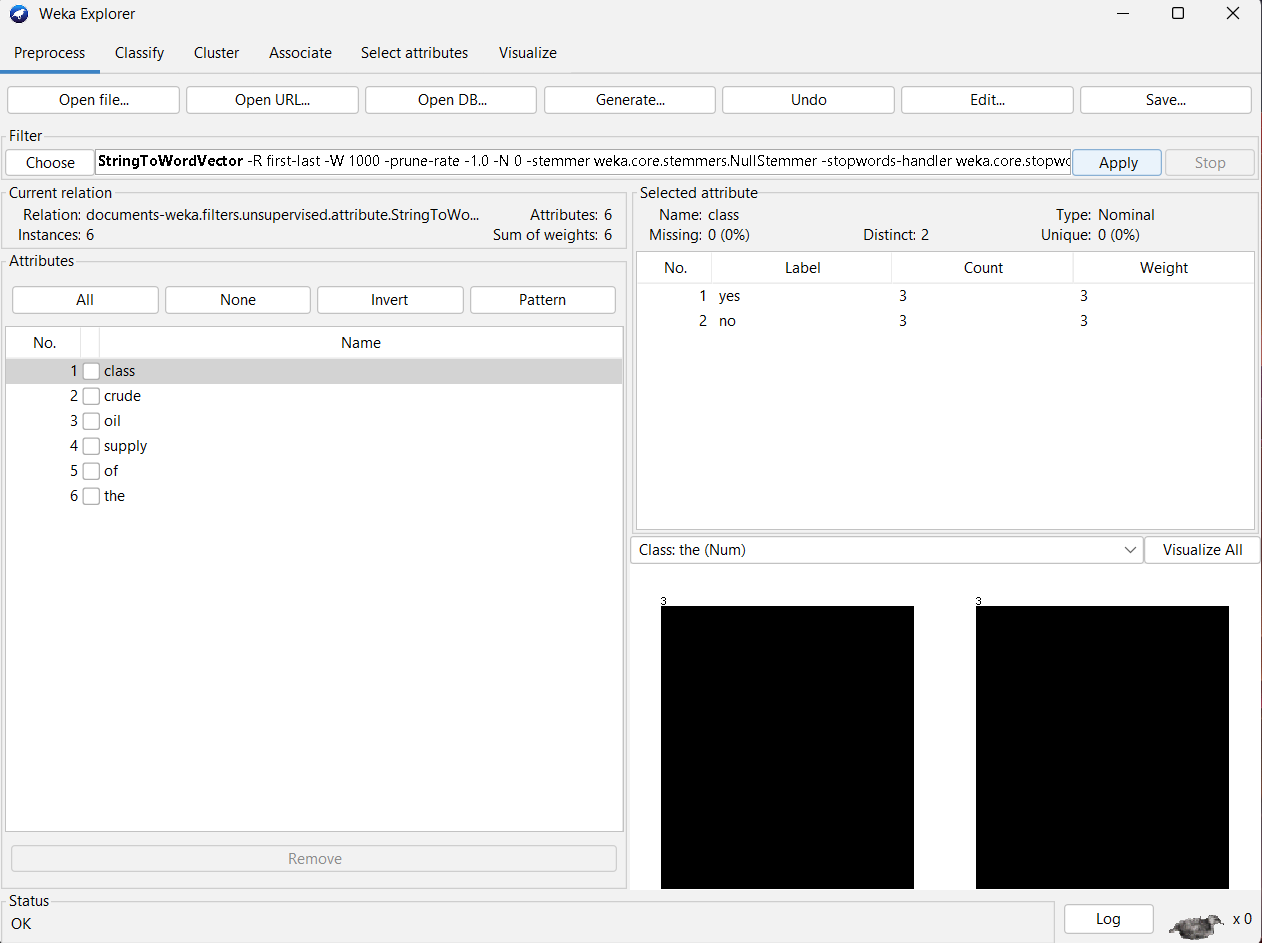
Description automatically generated

* The number of attributes generated by applying StringToWord with default option is 34.

A screenshot of a computer

Description automatically generated

* After change the value of the option minTermFreq to 2, only 5 attributes are generated.



**Exercise 2. Build a J48 decision tree from the last version of the data you generated.**

A screenshot of a computer

Description automatically generated

**Exercise 3. Classify the new documents in Table 2 based on the decision tree generated from the documents in Table 1. To apply the same filter to both training and test documents, use FilteredClassifier, specifying the StringToWordVector filter and J48 as the base classifier. Create an ARFF file from Table 2, using question marks for the missing class labels. Configure FilteredClassifier using default options for StringToWordVector and J48, and specify your new ARFF file as the test set. Make sure that you select Output predictions under More options in the Classify panel. Look at the model and the predictions it generates, and verify that they are consistent. What are the predictions?**

* Test documents:

A screenshot of a computer

Description automatically generated

* FilteredClassifier:

A screenshot of a computer

Description automatically generated

* Model:

A screenshot of a computer

Description automatically generated

* Prediction:

A screenshot of a computer program

Description automatically generated

|  |  |
| --- | --- |
| Documents | Classification |
| Oil platforms extract crude oil | yes |
| Canola oil is supposed to be healthy | no |
| Iraq has significant oil reserves | no |
| There are different types of cooking oil | no |