

Tomamos una muestra de las opiniones y se construye el archivo .arff para el programa weka el cual quedaría:

```
@RELATION text

@attribute textoDocumento string

@attribute docClass {false,true}

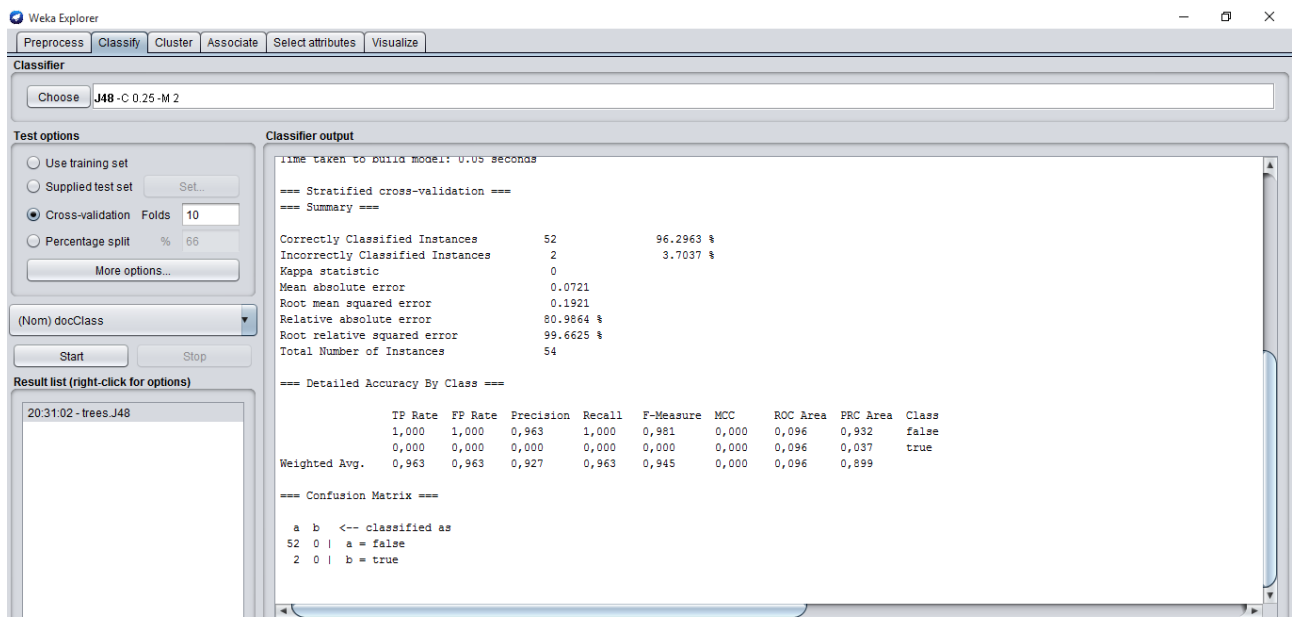
@data

" @bestnauniever: @malavemark @kerpen best news i heard day. @marcorubio are #praying you #fisen #debatenight ",false
" @msb971: i don't understand she be eligible run president #trumpsarmy #debatenight #podestâ€¦ ",false
" @meanpiastic: y'all too quick ôÿ", #debatenight ",false
" @msb971: i don't understand she be eligible run president #trumpsarmy #debatenight #podestâ€¦ ",false
" @kentkristensen1: i ran away was grabbing holding i never going vote #trump president #debatenight shameful respecô€ ",false
" @hillaryclinton: youâ€¢re proud hillary tonight. #debatenight #shewon ",false
" @mrntweet2: this how one my fb friends watched debate! #hereiamwithher #debatenight #debate #riggedmedia ",false
" @eunicesmadaula: #debates2016 #debatenight #debate mujeres violadas por bill clinton @realdonaldtrump ",false
" @hillaryclinton: youâ€¢re proud hillary tonight. #debatenight #shewon ",false
" @msb971: #obamacareinfourwords ôÿ't biggest pile shit #debatenight #trumpsarmy #draintheswamp #neverhillary @potus @barackoâ€¦ ",false
" @msb971: i don't understand she be eligible run president #trumpsarmy #debatenight #podestâ€¦ ",false
```

después abrir la aplicación y cargamos el archivo, despues seleccionamos el filtro **StringToWordVector** para formar n -gramas de tamaño 1 y 2

The screenshot shows the Weka Explorer application. The 'Filter' tab is active, and the 'StringToWordVector' filter is selected with parameters: -R first-last -W 1000 -prune-rate -1.0 -N 0 -stemmer weka.core.stemmers.NullStemmer -stopwords-handler weka.core.stopwords.Null -M 1 -tokenizer weka.core.tokenizers.NGramTokenizer -max 2 -r. The 'Current relation' section shows 'Relation: text-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-r...' with 594 attributes and 54 instances. The 'Attributes' list on the left shows various attributes, including '30 years', '@annwelterpayne', '@antoniodelotero', '@awkiwardpuppets', '@barackoâ€¦', '@bestnauniever', '@blackpphines', '@cedpaofficial', '@chescaleigh', and '@chescaleigh the'. The 'Selected attribute' section shows 'Name: docClass', 'Missing: 0 (0%)', 'Distinct: 2', and 'Type: Nominal'. A table below shows the distribution of the 'docClass' attribute: '1 false' with a count of 52 and weight of 52.0, and '2 true' with a count of 2 and weight of 2.0. The 'Class: docClass (Nom)' is selected, and a 'Visualize All' button is visible. A small bar chart at the bottom shows the distribution of the 'docClass' attribute, with a blue bar for 'false' (52) and a red bar for 'true' (2).

despues seleccionamos la pestaña de classify que nos dara la opcion de escoger el algoritmo J48 que esta basado en ID3 los cuales son los arboles de desicion



tenemos el árbol que construyo el algoritmo, para este caso tenemos que solo selecciono a “FALSE”

false (54.0/2.0)

tenemos la gráfica de como se comporta la clasificación

