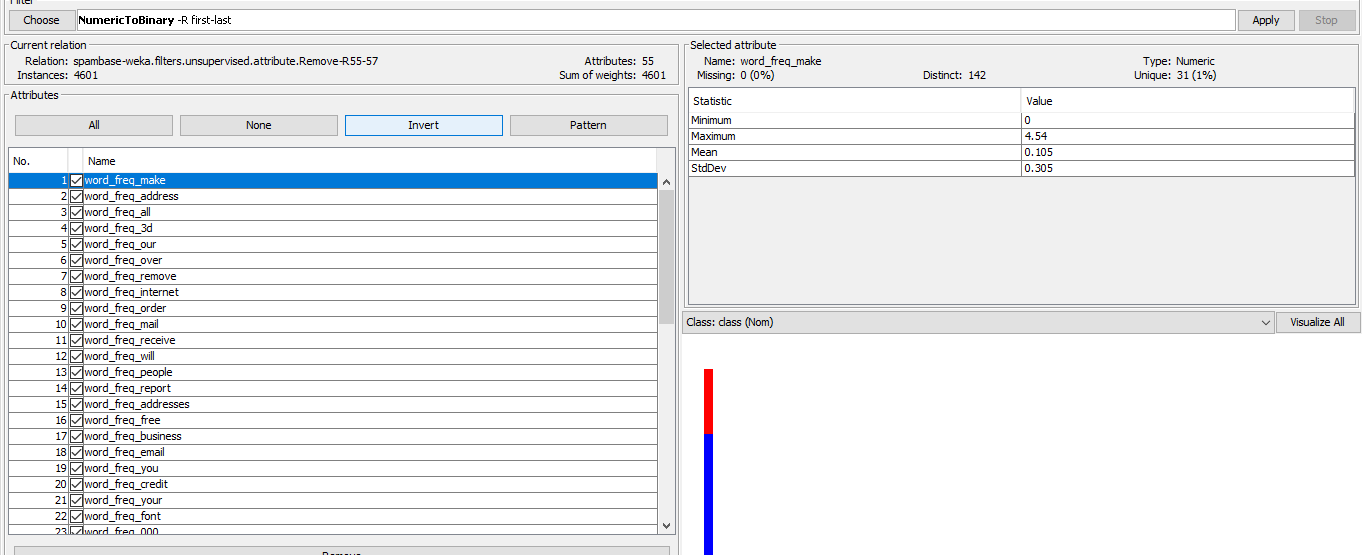
Lab 4 Home Task

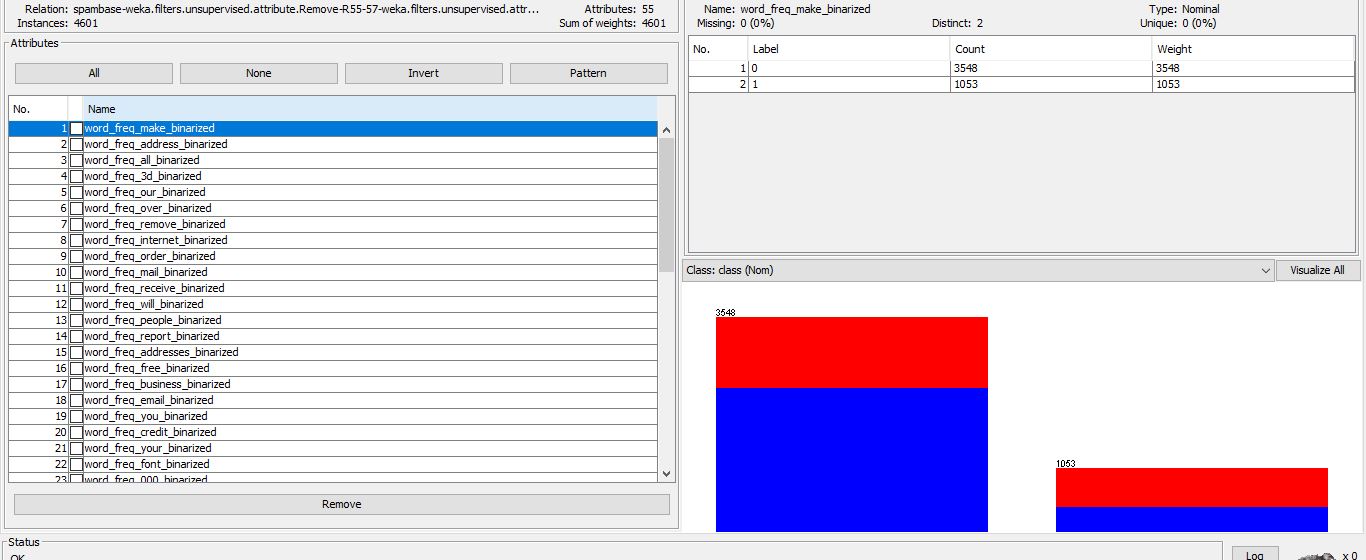
Removed columns

* capital\_run\_length\_average
* capital\_run\_length\_longest
* capital\_run\_length\_total attributes.

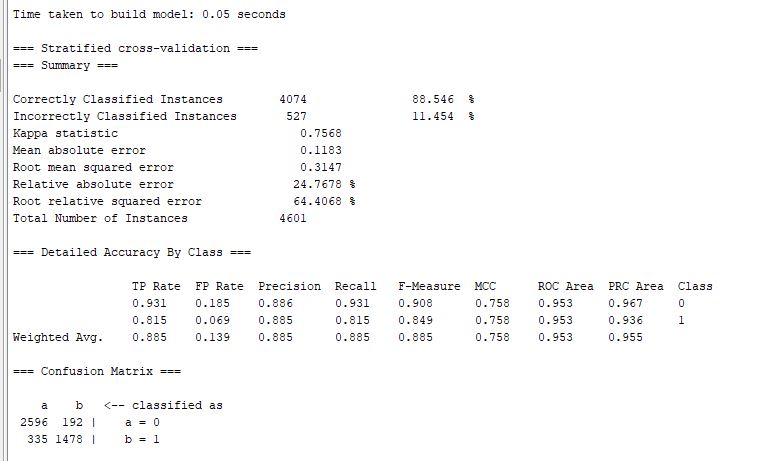


Bag of words.

Applying NumericToBinary filter, data is converted into Boolean. This makes it bag of words.



Applying Naïve bayes classifier on this using default parameters.

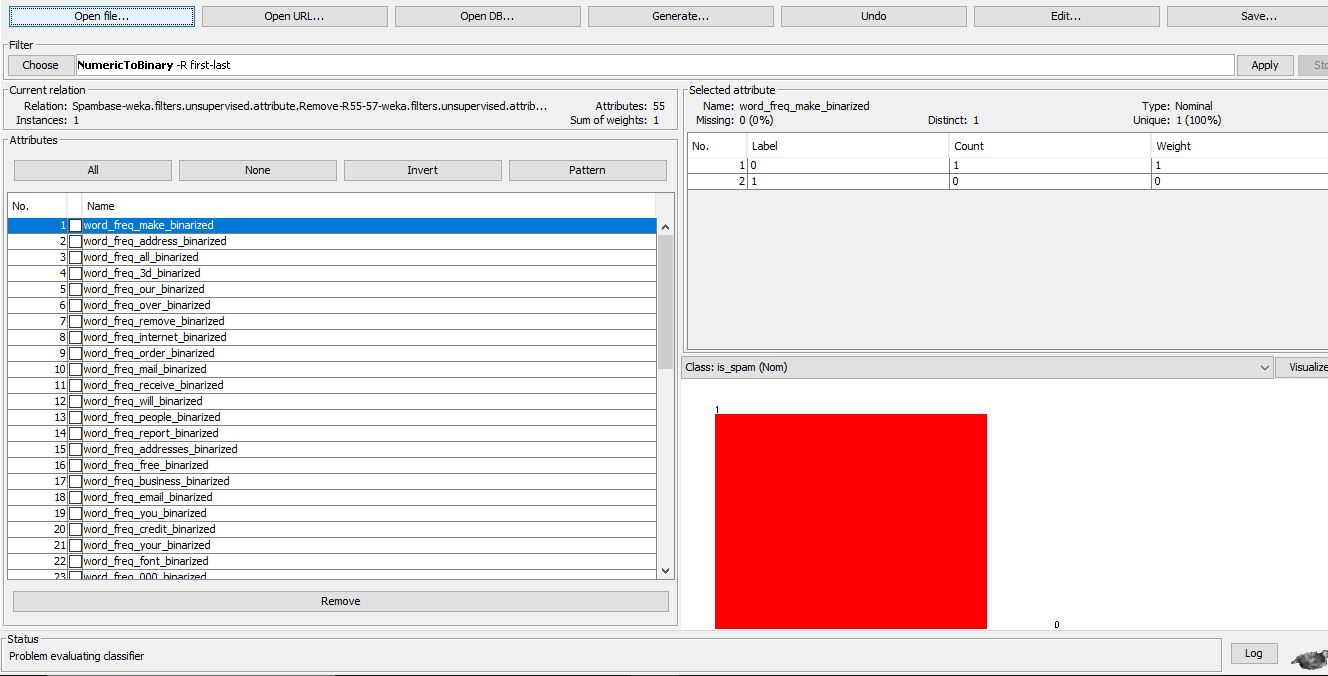


When the assumption of feature independence holds reasonably well—that is, when features are conditionally independent given the class label—and when datasets with a comparatively small number of features are used, Naive Bayes classifiers often perform well. This dataset appears to support the assumption to a reasonable degree, which would result in good performance. Additionally, Naive Bayes may result in subpar performance if the dataset's feature independence requirement is not met. The classifier took 0.05 seconds to construct, and its computational efficiency allows it to be scaled to huge datasets.

P (3d | spam) = 4 / 5 = 0.8

P (3d | non-spam) = 3 / 5 = 0.6

These probabilities represent the likelihood of observing the word "3d" given an email is classified as spam or non-spam.  
  
Now I have opened test data provided and will be using to train our classifier and apply the learnt classifier to a dedicated test set.



test data uploaded in weka

applying naïve bayes on it,

Result is after supplied test set as a parameter.

