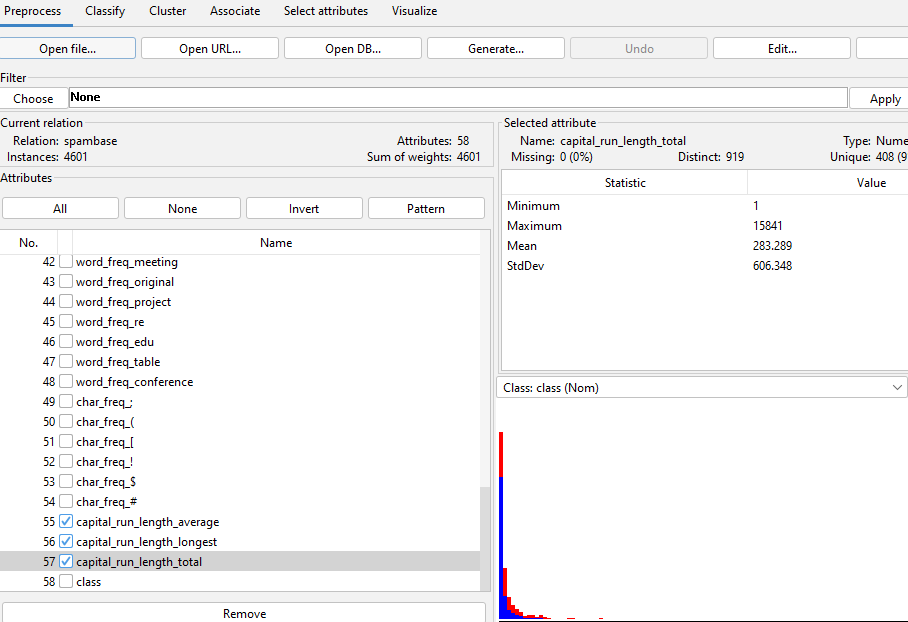
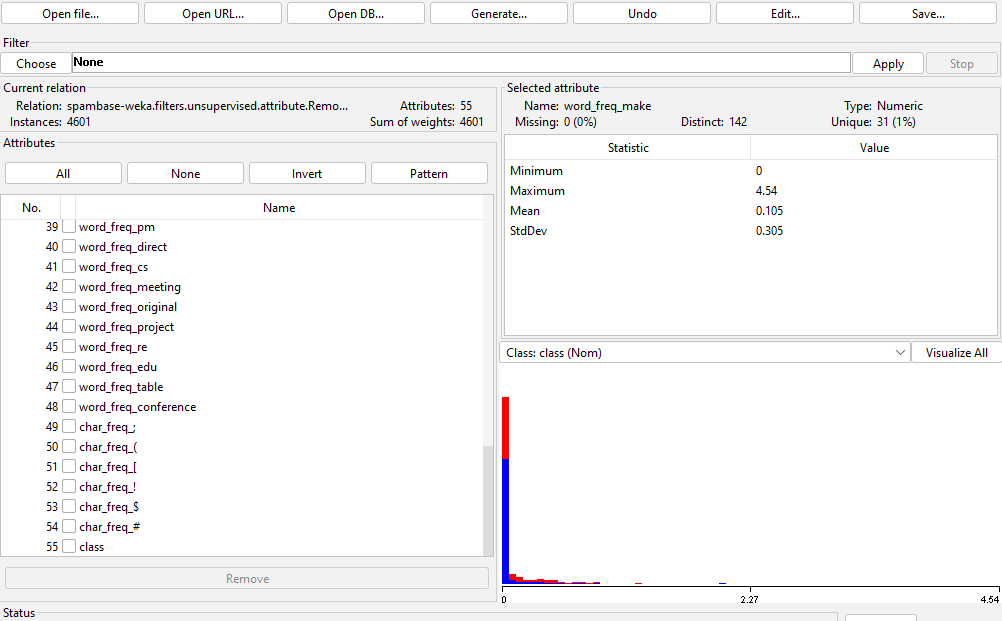
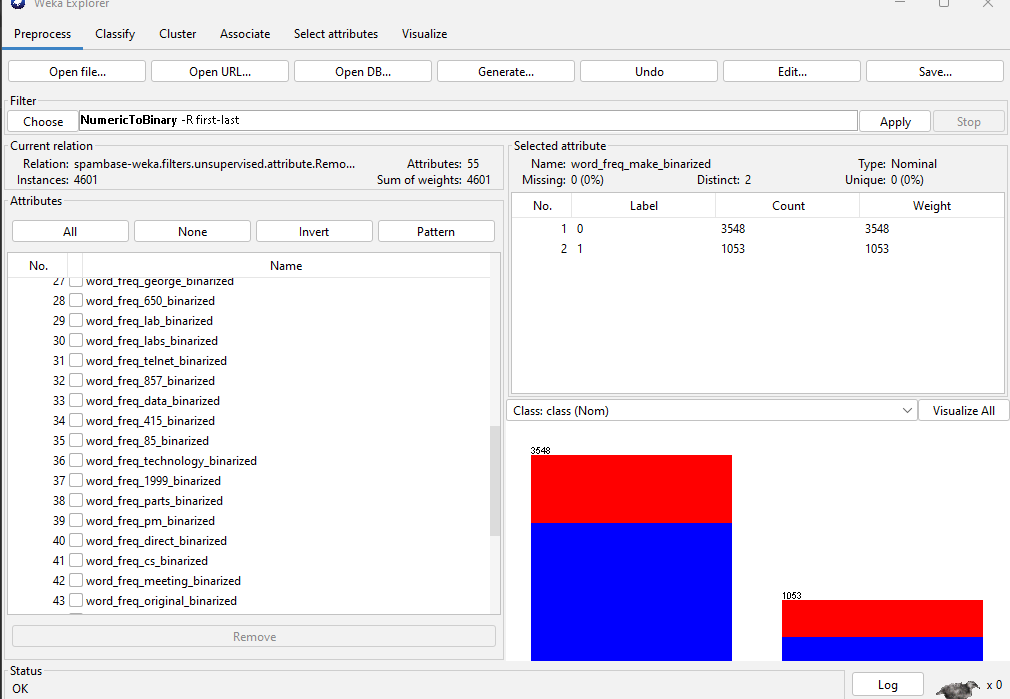
**Naïve Bayes**

* **Removing Attributes:**

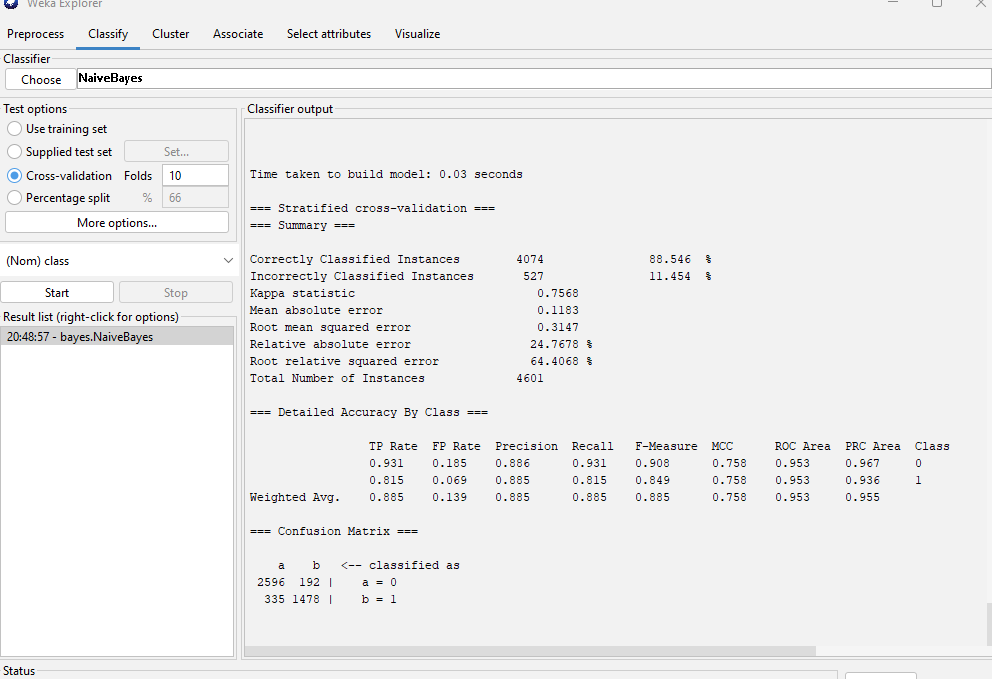




* **Converting to Binary:**



* **Applying Naïve Bayes:**



* **Reason for Good Performance:**

The data is linearly separated and follows a specific pattern making the work easier for the classifier.

* **Practical Problems:**

The practical problems we could face include overfitting, inefficient models, not understanding the dataset, etc.

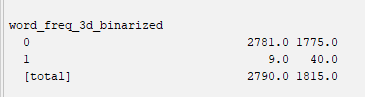
* **Time Taken:**

The classifier took 0.03 seconds to train and classify the dataset.



The time taken identifies that naïve Bayes is highly scalable for large datasets. Naïve Bayes is said to be highly scalable because of its simplicity and independence. It works by calculating conditional probabilities which makes it optimized.

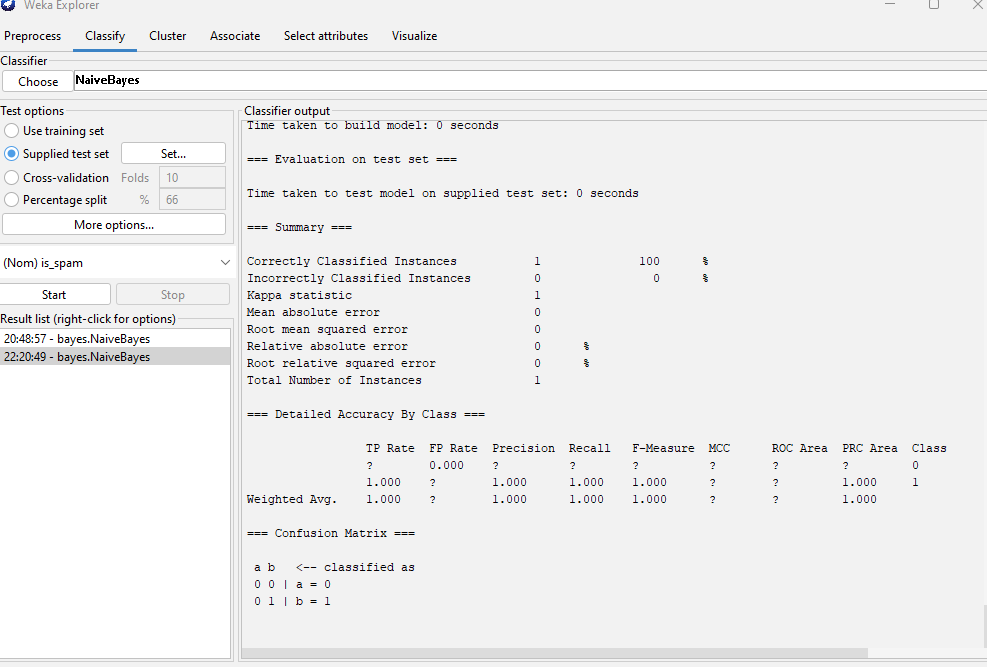
* **Spam/Not Spam(word: 3d):**



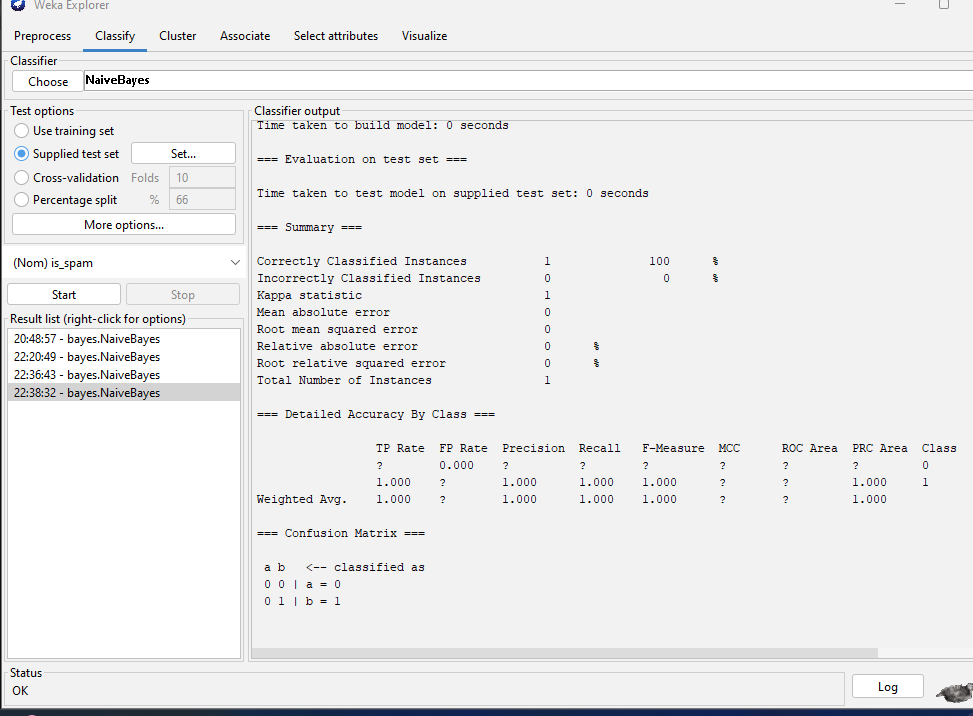
Class 0 (Not Spam): 1755 instances

Class1(Spam): 40

**Test**



* The classifier correctly classifies spam email. The precision, recall, and F1-score metrics for “spam” indicate perfect performance, suggesting that the classifier accurately identified all instances of spam. This is also confirmed by the confusion matrix.
* **Classifier on Modifed Test Data:**



As the precision, recall, and F1-score for class 1 (spam) are all perfect, indicating that all instances of spam were correctly identified so the results have not changed. They were correctly identified as “spam” before as well .