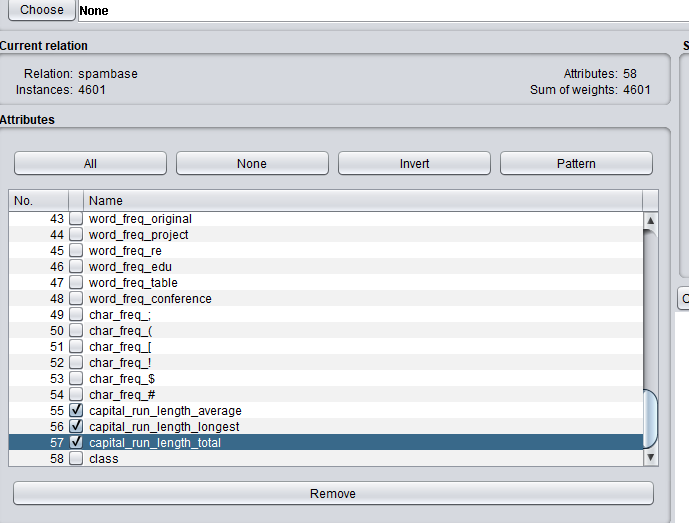
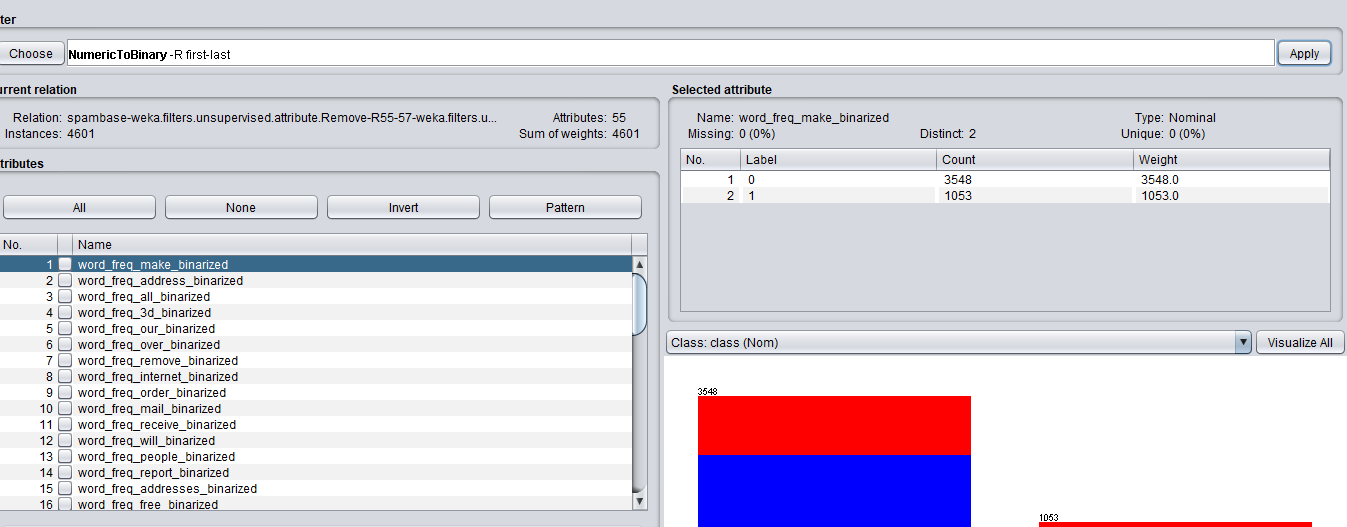
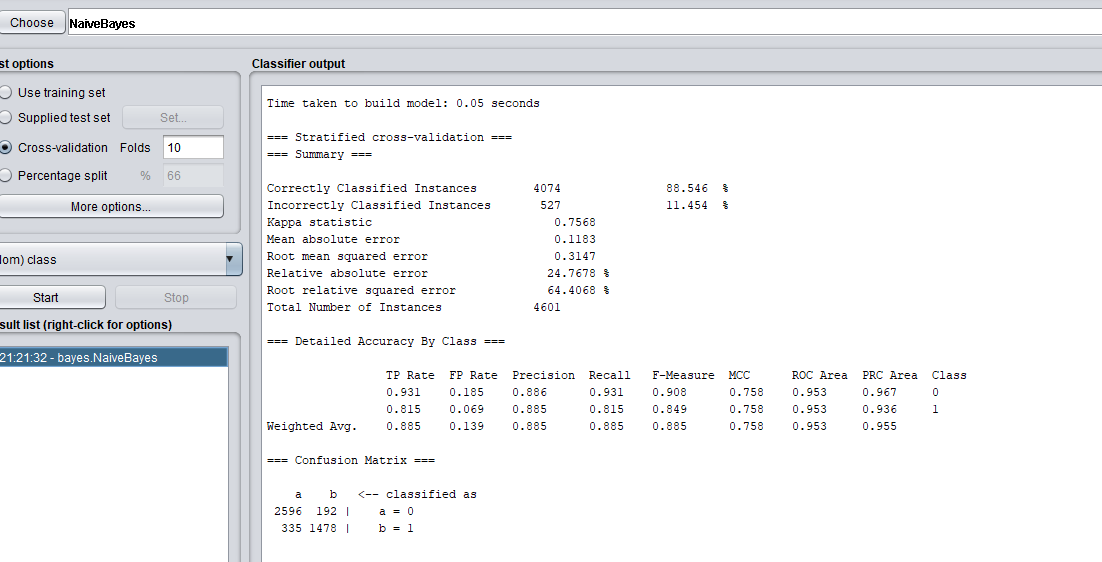
Lab 4:

1. Preprocess the file and remove the attributes



After removing the attribute, I apply numerictobinary for msgs



The correctly classified instances is 88% and incorrectly classified is 11 

* The Naïve Bayes classifier's good performance is due to its simplicity, scalability, and effectiveness with text data. Making the independence assumption simplifies the model and reduces computational complexity, allowing for efficient processing of high-dimensional feature spaces. Without this assumption, the main practical problems would include high dimensionality, data sparsity, the curse of dimensionality, and overfitting.
* Naïve Bayes classifiers are generally considered to be quite scalable to large datasets, especially those with high-dimensional feature spaces, like text data. And my dataset takes a few seconds to train my data.

1. 1)

* In Naïve Bayes classification, the probability of an email belonging to a class (e.g., spam or not spam) is computed using Bayes' theorem:

P(Class|Attributes) = (P(Class) \* P(Attributes|Class)) / P(Attributes)

Therefore, the conditional probabilities are:

P(3d|spam) = (1 + 1) / (1 + 1 + 1 + 1) = 2 / 4 = 0.5

P(3d|non-spam) = (1 + 1) / (1 + 1 + 1 + 1) = 2 / 4 = 0.5

So, according to the Naïve Bayes model, the probability of observing the word "3d" given that an email is spam is 0.5, and the probability of observing the word "3d" given that an email is non-spam is also 0.5.

2) A screenshot of a computer

Description automatically generated

Now, we can compute the probabilities:

P(3d|spam) = (1) / (2 + 1) = 1/3 ≈ 0.3333

P(3d|non-spam) = (3) / (3 + 3) = 3/6 = 1/2 = 0.5

3)

A screenshot of a computer

Description automatically generated

Now, we can compute the probabilities:

P(3d|spam) = (3) / (2 + 1) = 3/3 = 1

P(3d|non-spam) = (3) / (3 + 3) = 3/6 = 1/2 = 0.5

So, according to the Naïve Bayes model, the probability of observing the word "3d" given that an email is spam is 1, and the probability of observing the word "3d" given that an email is non-spam is 0.5.

