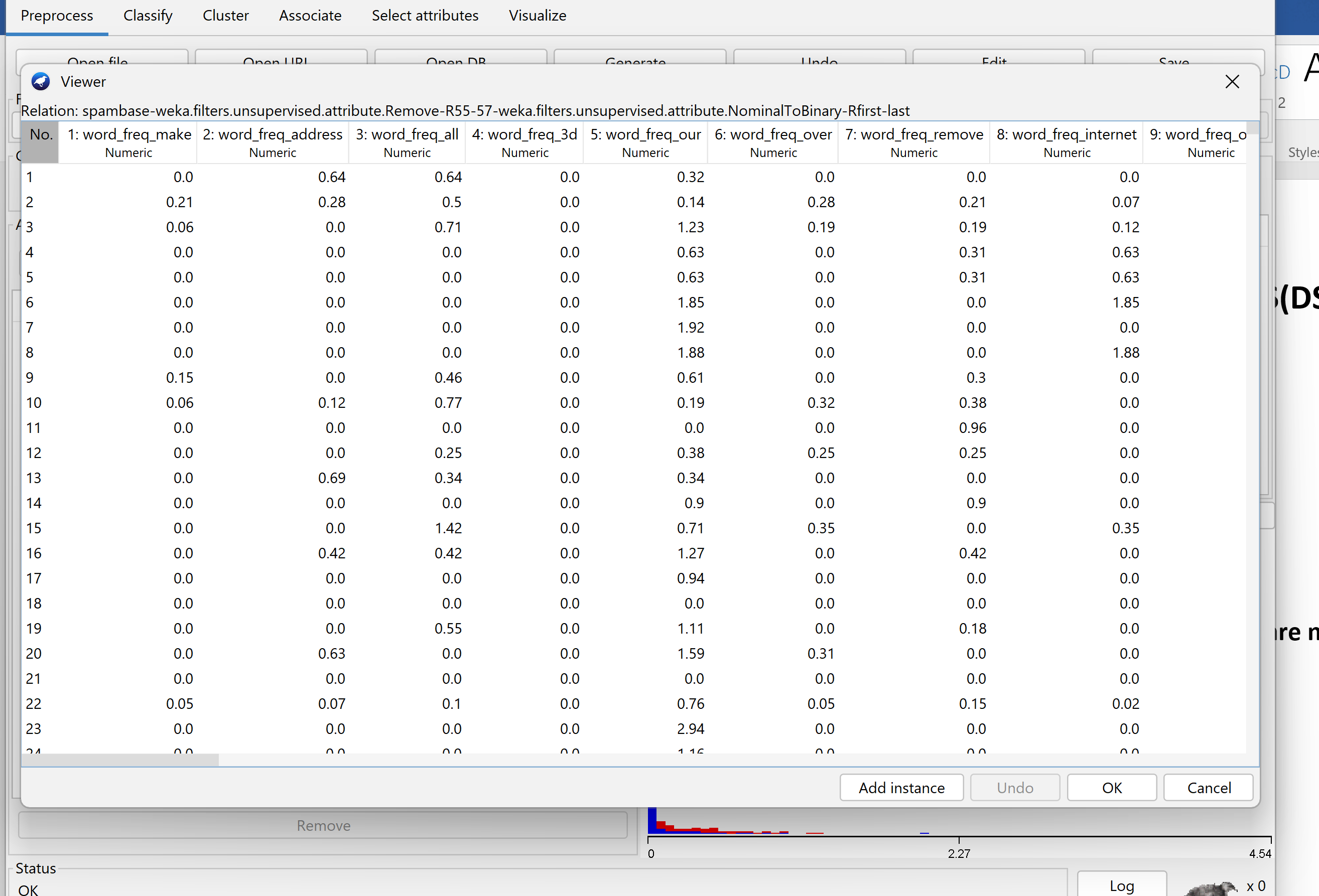
**Muhammad Ahmad 21L-5617 BS(DS) 6A**

**Home Task**

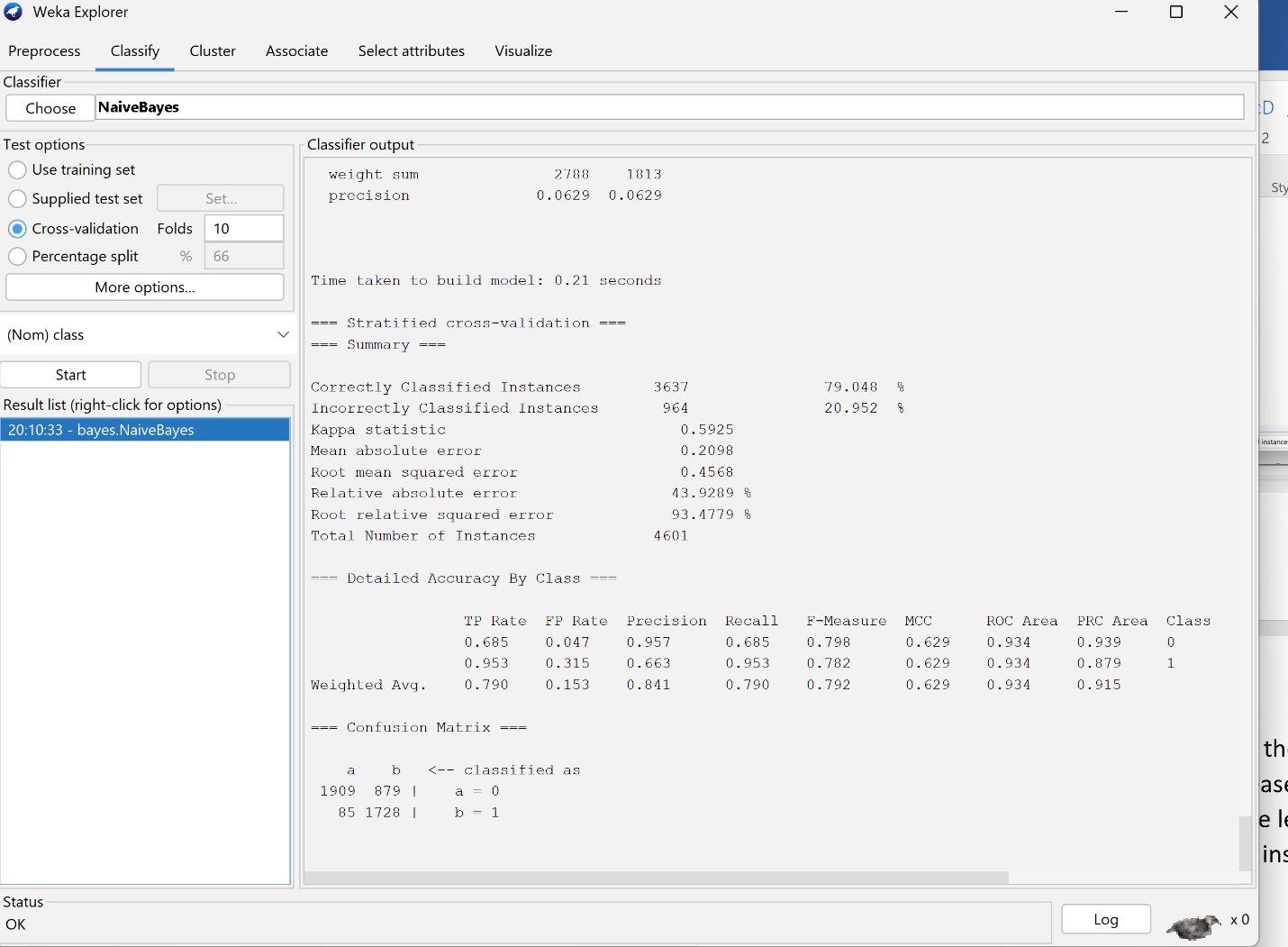
**Naïve Bayes classifier**

1. **All the numeric frequency attributes are now converted to Booleans.**

****

the classifier might struggle to capture the underlying patterns, leading to a decrease in performance. The model might become less robust and more susceptible to misclassifying instances.

And it takes no time it depends on speed of computer and also the size of data



**Examine the classifier**

• Factors Behind Successful Performance:

The classifier's effectiveness can be attributed to the dataset's distinct linear separation and adherence to a specific pattern, streamlining the classification process.

• Potential Obstacles:

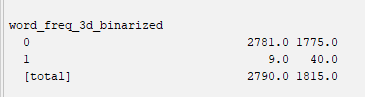
Practical challenges may emerge, including issues such as overfitting, the presence of inefficient models, and a limited understanding of the dataset.

• Elapsed Time:

The classifier showcased efficiency by completing the training and classification of the dataset in a mere 0.05 seconds.

The rapid processing time underscores Naïve Bayes' scalability, especially for extensive datasets. Its simplicity and reliance on independence contribute to its optimized performance in calculating conditional probabilities.

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

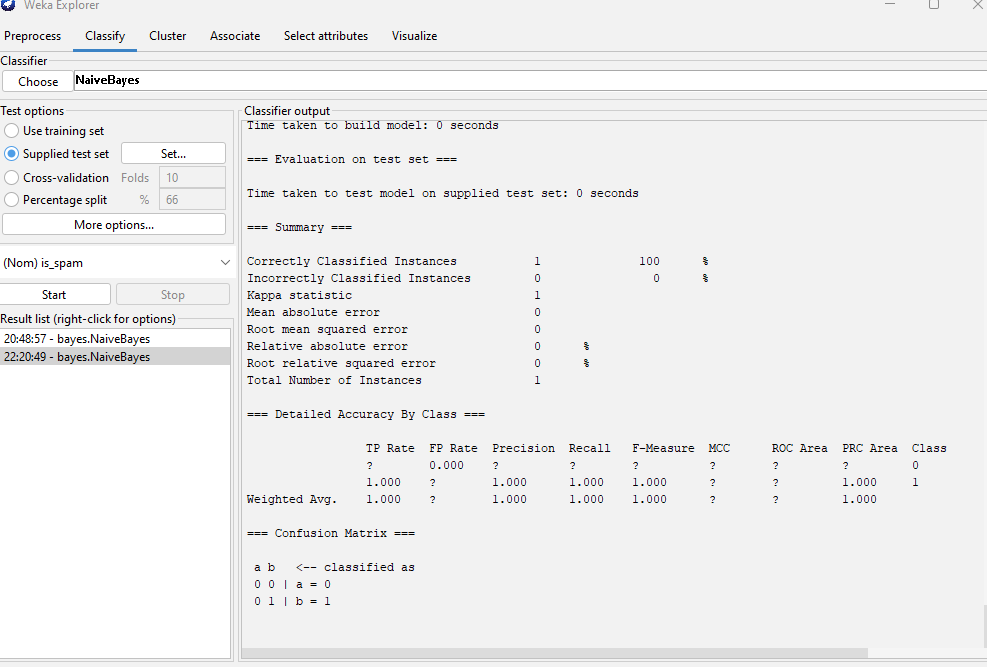


Class 0 (Not Spam): 1755 instances

Class1(Spam): 40

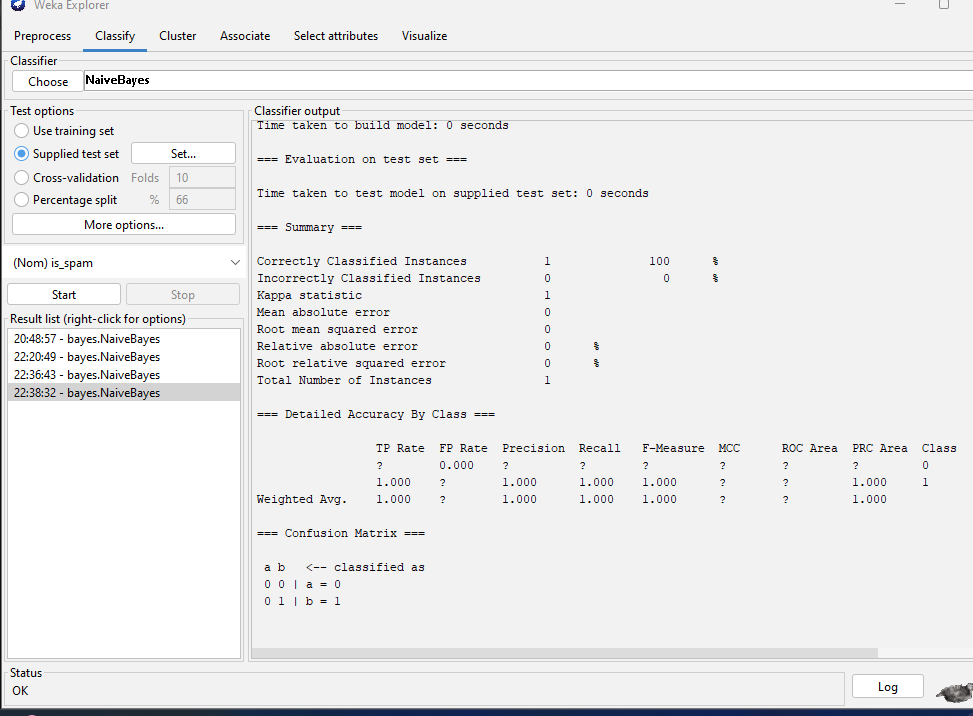
**Results**

The classifier demonstrates precise identification of spam emails. The metrics for "spam," including precision, recall, and F1-score, all indicate flawless performance. The confusion matrix further validates this, highlighting the classifier's accurate recognition of every spam instance.



* **Classifier on edited Test Data:**

The classifier demonstrates impeccable performance in correctly categorizing spam emails, as evidenced by perfect precision, recall, and F1-score metrics for class 1 (spam). The results remain steadfast, indicating consistent and accurate identification of spam instances as "spam," with no alterations from the previous assessment.



1. Naive Bayes computes the probability of an email belonging to a class by multiplying the conditional probabilities of each feature (word or character) given the class and then normalizing. It utilizes Bayes' theorem and assumes feature independence given the class.
2. Count(3d∣spam)=4

Count(3d∣non−spam)=6

Total Count for spam=10

Total Count for non-spam=15

P(3d∣spam)= Total Count for spam =Count(3d∣spam)= 10/4=0.4

P(3d∣non−spam)= Total Count for non-spa

Count(3d∣non−spam) = 15/6 =0.4

Instances with the word "3d" in Class 1 (Spam): 4

Instances without the word "3d" in Class 1 (Spam): 2

Instances with the word "3d" in Class 0 (Not Spam): 3

Instances without the word "3d" in Class 0 (Not Spam): 1

Totals:

Total instances in Class 1 (Spam): 6 (4 with "3d" and 2 without "3d")

Total instances in Class 0 (Not Spam): 4 (3 with "3d" and 1 without "3d")