**AMERICAN INTERNATIONAL**A close up of a sign

Description automatically generated

**UNIVERSITY-BANGLADESH**

**Assignment-Mid Term**

**NAME: MD. AKHTARUZZAMAN EMON**

**ID: 18-36074-1**

**COURSE: DATA WAREHOUSING AND DATA MINING**

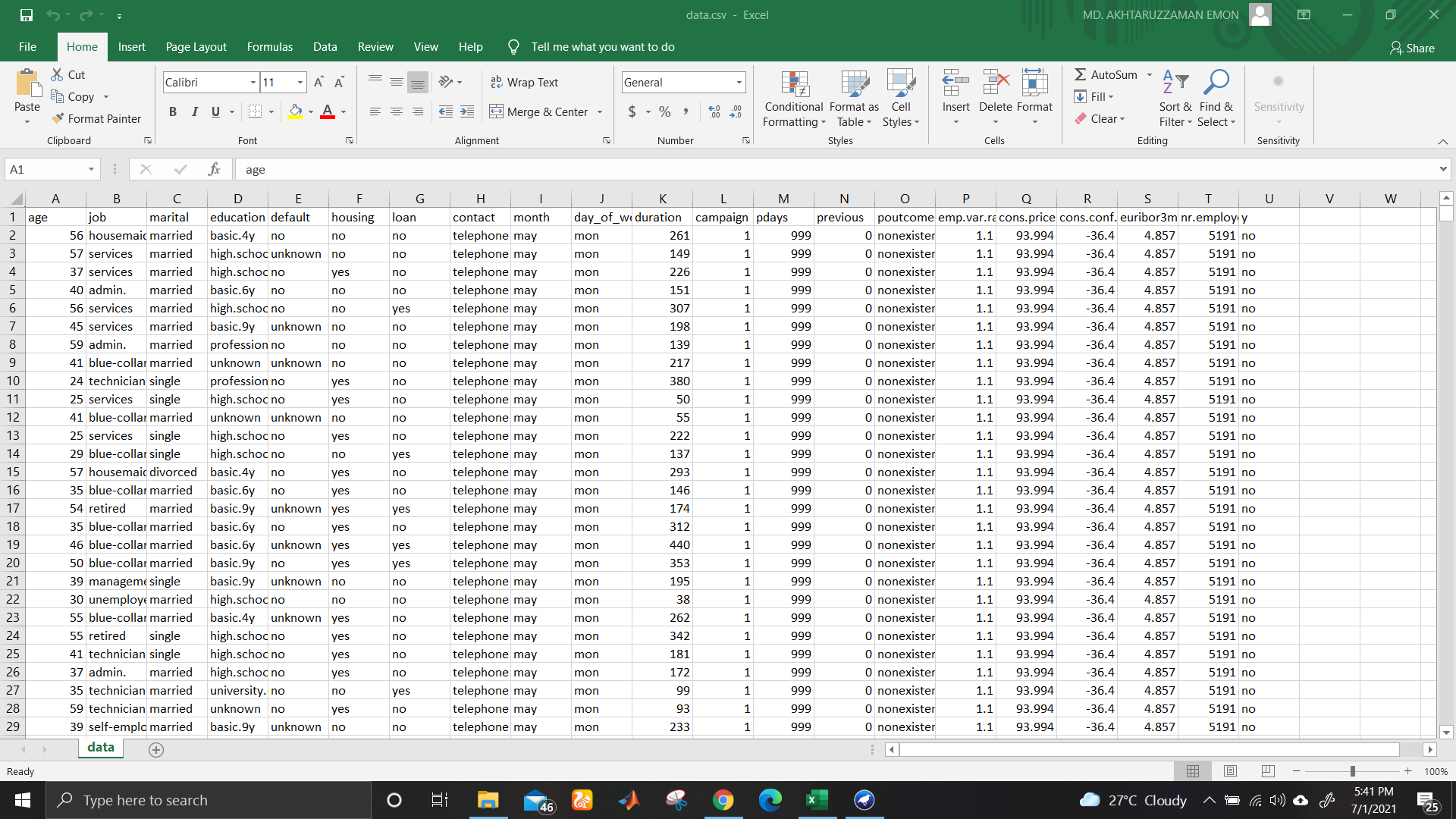
**SECTION: A**

**SUBMITTED TO- DR. MD. MAHBUB CHOWDHURY MISHU**

**Introduction:**

A data mining function called classification allocates objects in a collection to desired groups or classes. Classification's purpose is to correctly anticipate the target class for each case in the data. In this report I used K-Nearest Neighbor (KNN) procedure. The K-Nearest Neighbor algorithm is an example of a "lazy learner," meaning it does not build a model using the training set until the data set is queried. For this report I used a data set of a bank. There are many data of people in this data set. We use Weka software to find the result.

**Data Set:**

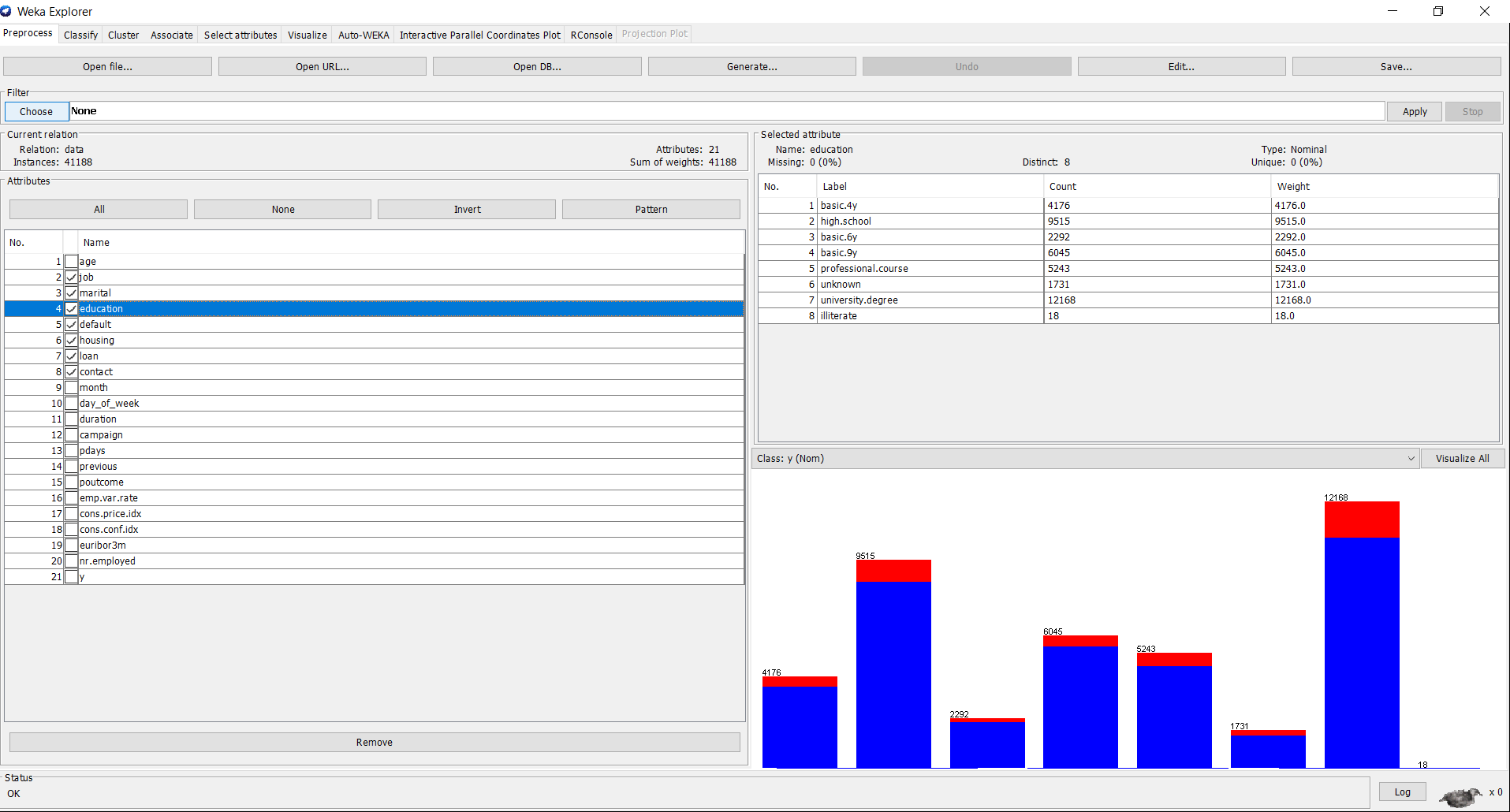


In this data set there are 21 columns and 41189 rows.

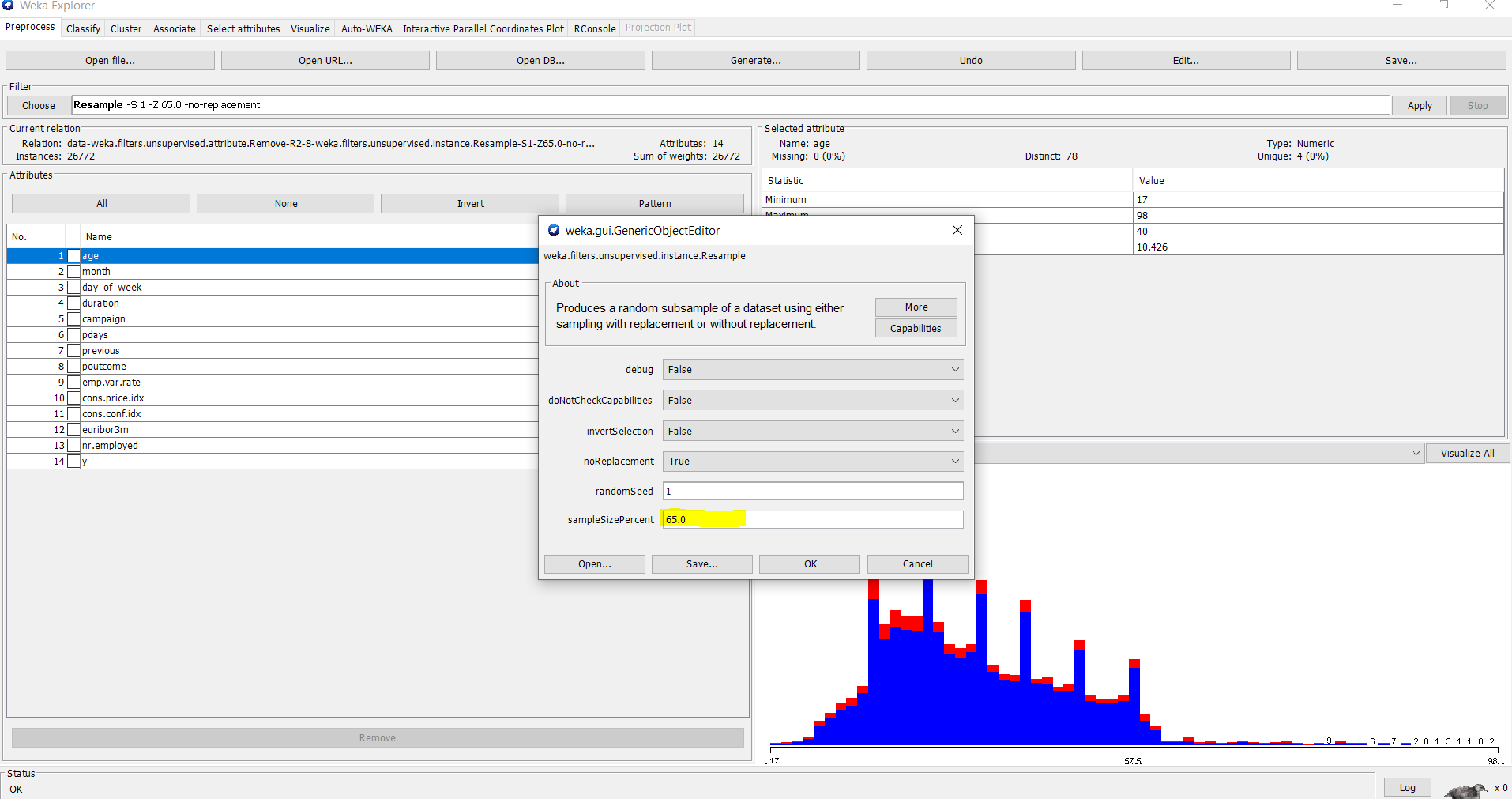
Dataset reference: <https://www.kaggle.com/brijbhushannanda1979/bank-data>

**Procedure:**

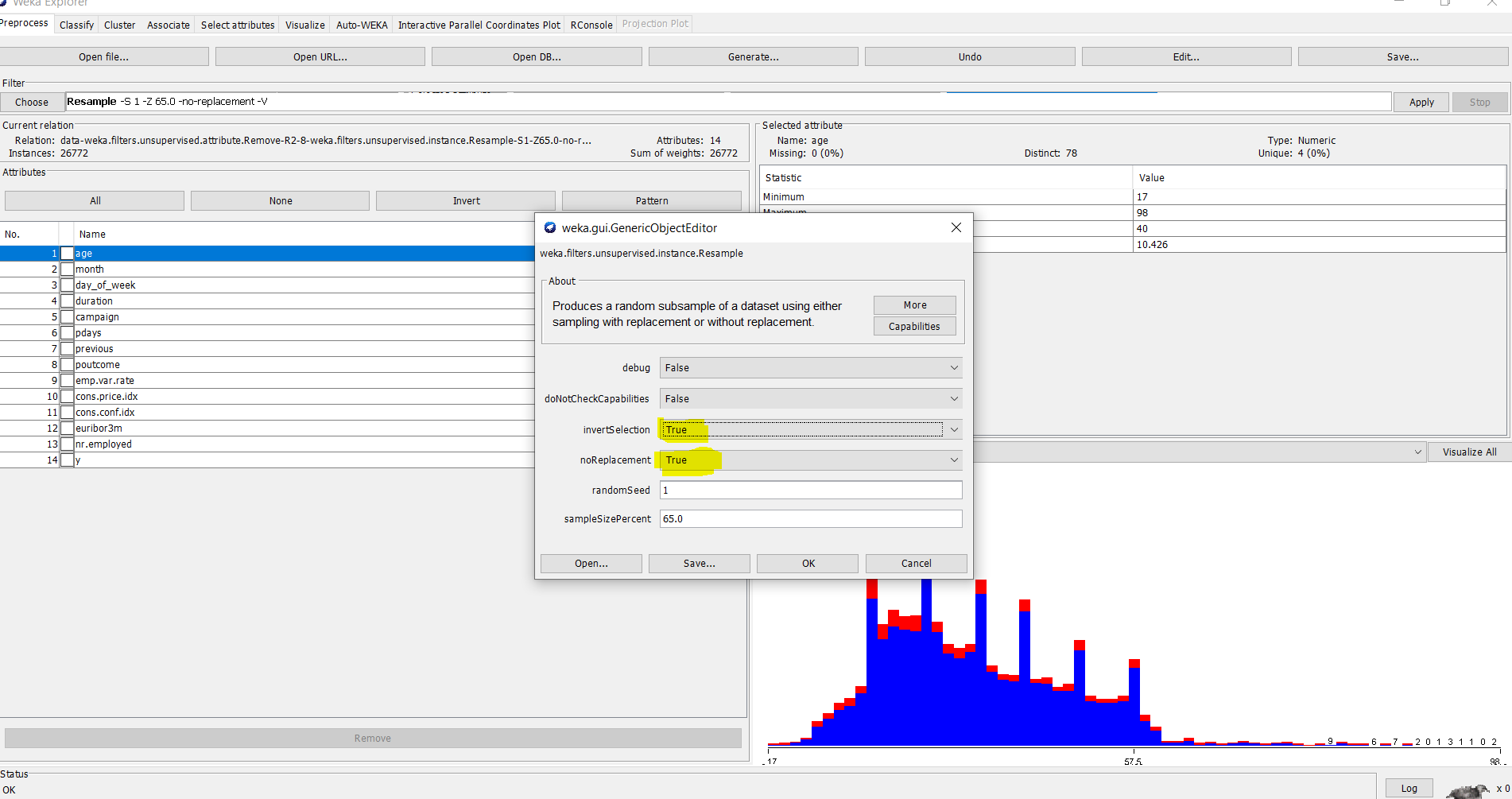
At first, we need to open the data set file in Weka software. Then we need delete unnecessary data from the data set. Then we have to follow the algorithms of K-Nearest Neighbor algorithm, Naïve Bayes.



Then we have to set the data in two set which are training set and test set.



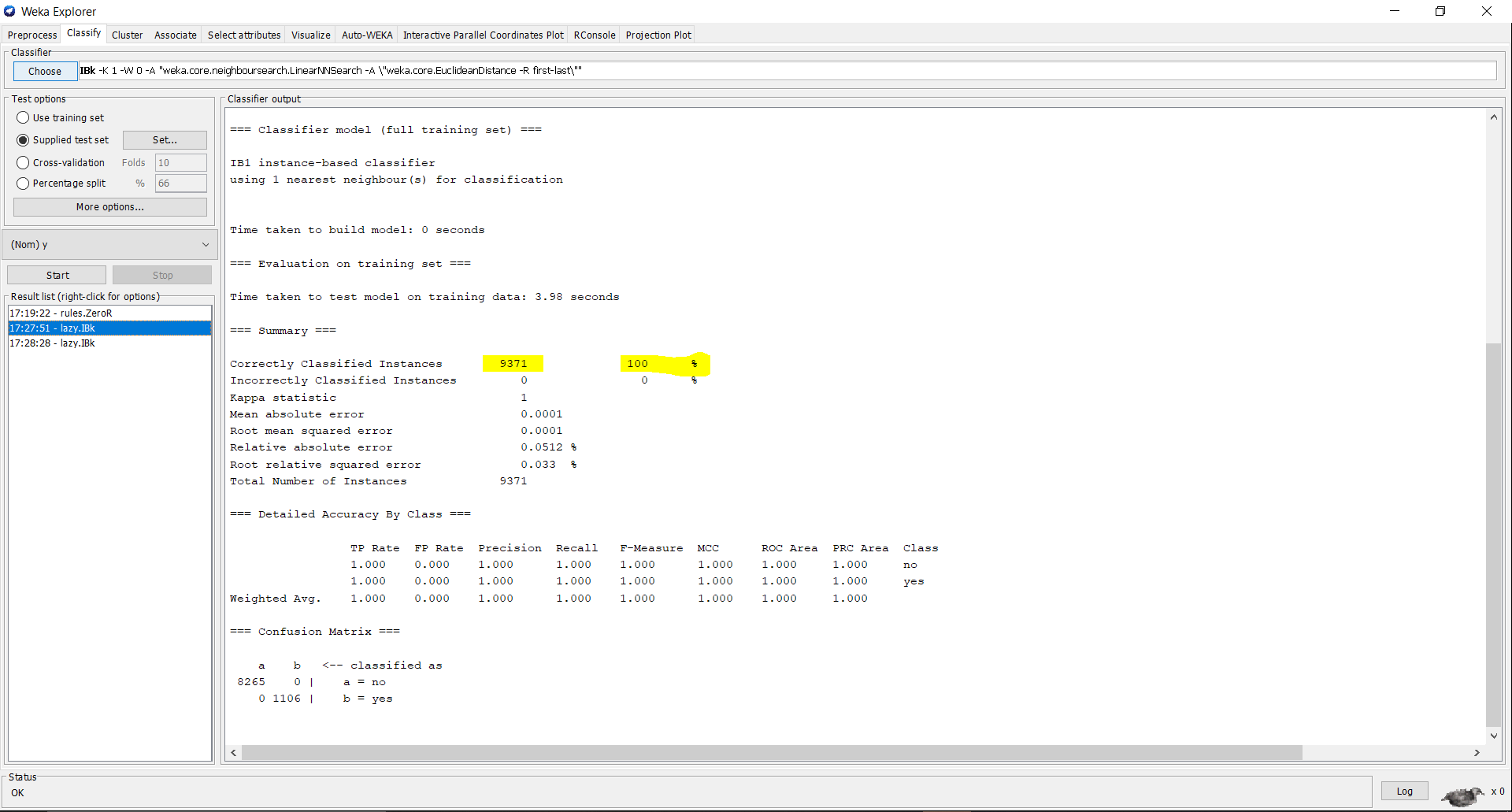
Here we set 65% of data for training set.



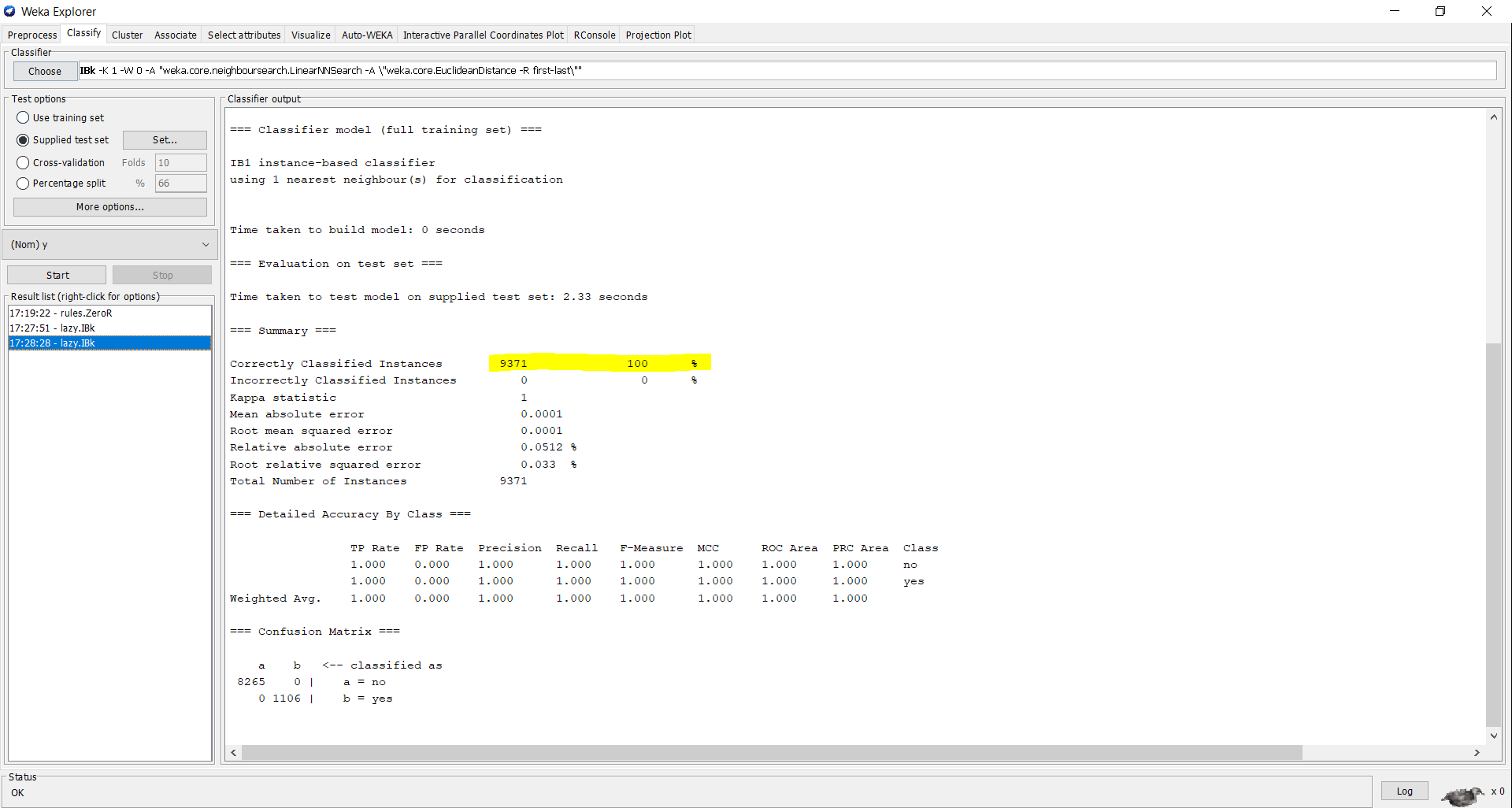
Here we set 55% of data for test set.

**Result of K-Nearest Neighbor (KNN):**

Training set:

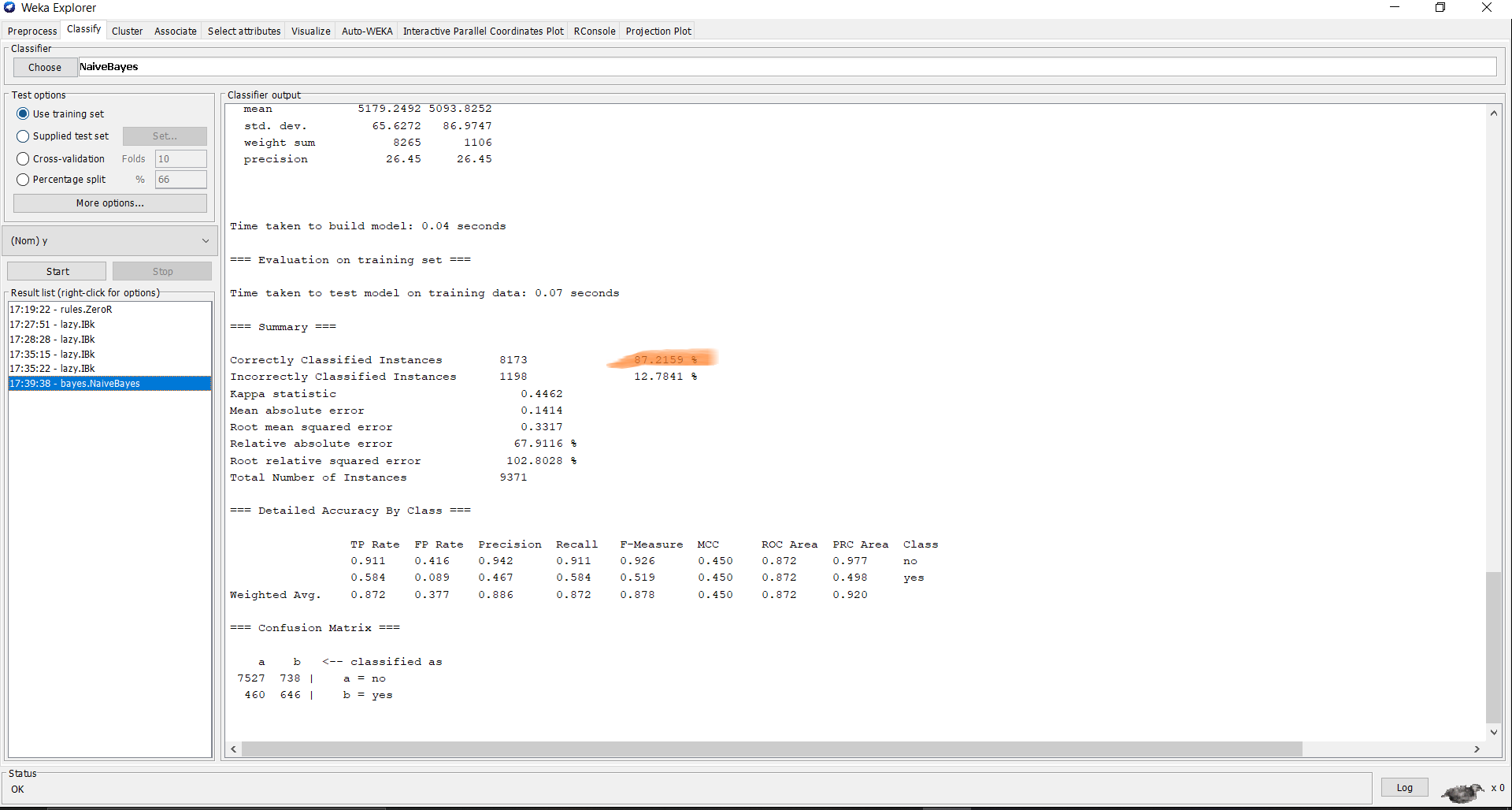


Test set:

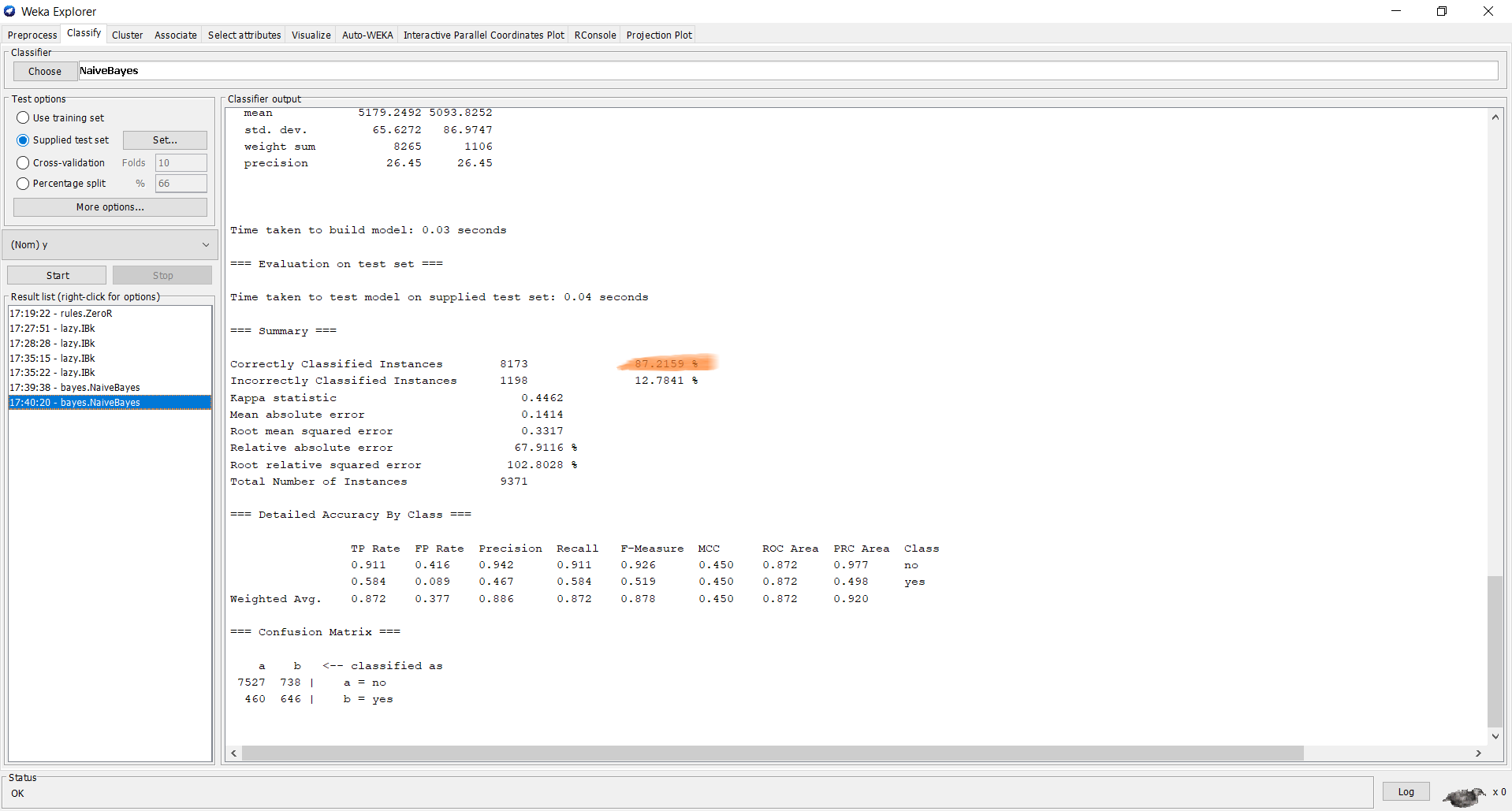


**Result of Naïve Bayes:**

Training set:



Test set:



**Results:**

By using these particular algorithms, we got some different results. We used K-Nearest Neighbor algorithm, Naïve Bayes procedure. From these K-Nearest Neighbor (KNN) procedure is the best algorithm for this problem. I got 100% accuracy of KNN in training set and also got 100% in test set. On the other hand, in Naïve Bayes procedure I got 87.22% accuracy in training set and test set. So, I prefer the K-Nearest Neighbor procedure.