**[DATA WAREHOUSING AND DATA MINING](https://portal.aiub.edu/Student/Section?q=shHyi79w9kTy%2FR%2BmvDK8ZQ%3D%3D)**

**SEMESTER:** Fall 2021-2022

**FINAL-TERM ASSIGNMEN**

**SUBMITTED BY:**

**STUDENT NAME**: Kainath, MD Tazul Islam

**STUDENT ID:** 19-39341-1

**SECTION:** D

**DEPARTMENT:** CSE

**SUBMITTED TO:**

**COURSE TEACHER:** Tohedul Islam

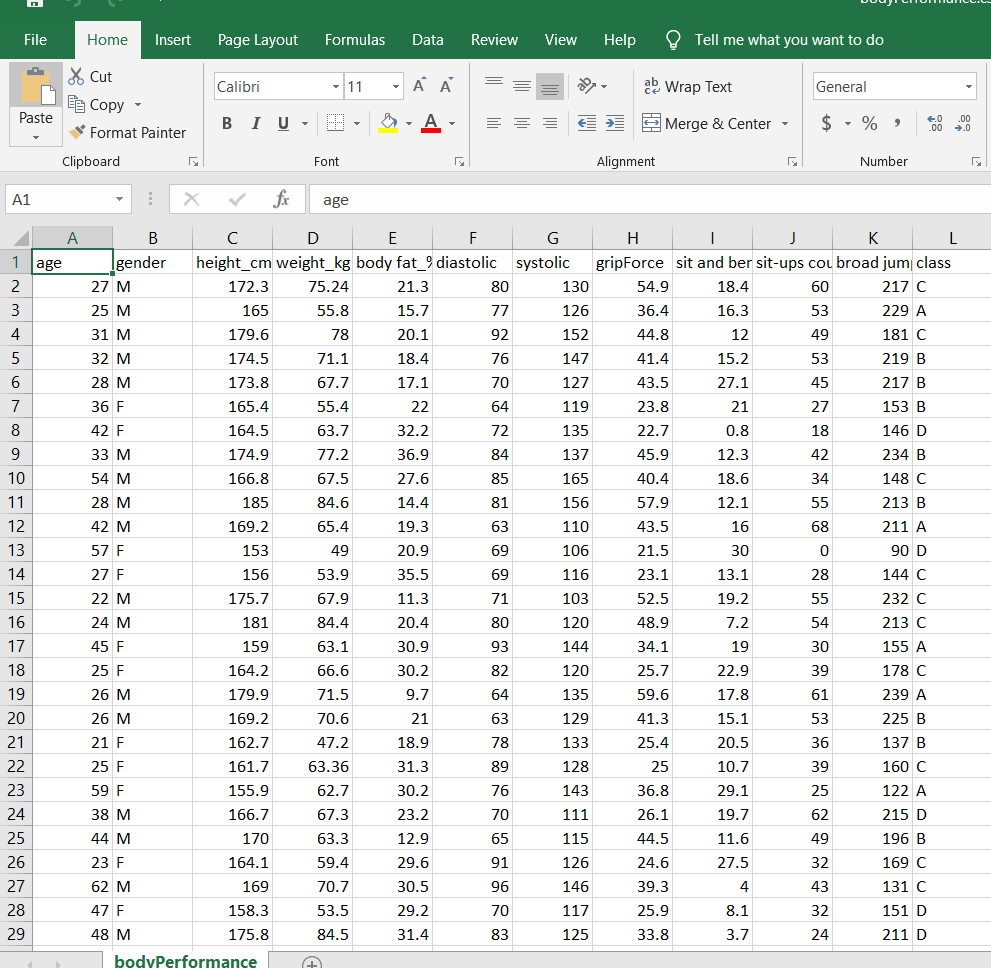
**Task-1**

**Introduction**

Body performance Data is the dataset that I used to apply classify. The information is structured in such a way that it may be used to certify the grade of performance based on age and certain exercise performance data. The information was gathered from the dataset area of kaggle. It's a categorization with many levels.

The collection has 349 instances with a total of 12 characteristics (Attributes).

Dataset:

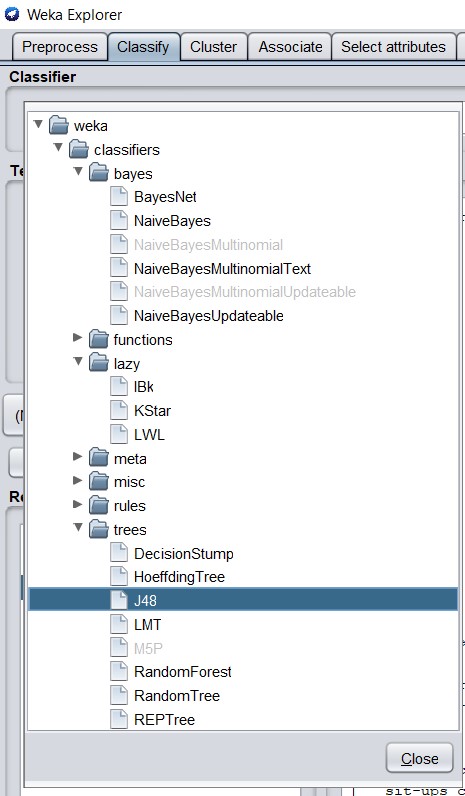


The dataset's first 29 instances (out of 349) are shown in the attached image. As an example of the datasets, please see the attachment.

Attributes:

* age : 20 ~64
* gender : F,M
* height\_cm
* weight\_kg
* body fat\_%
* diastolic : diastolic blood pressure (min)
* systolic : systolic blood pressure (min)
* gripForce
* sit and bend forward\_cm
* sit-ups counts
* broad jump\_cm
* class : A,B,C,D ( A: best) / stratified

The following step will be to select categorize from Weka. Weka is a software package that includes visualization tools and algorithms for data analysis and predictive modeling. We will utilize Weka to analyze data for this task.



I chose three classifiers from the given options.

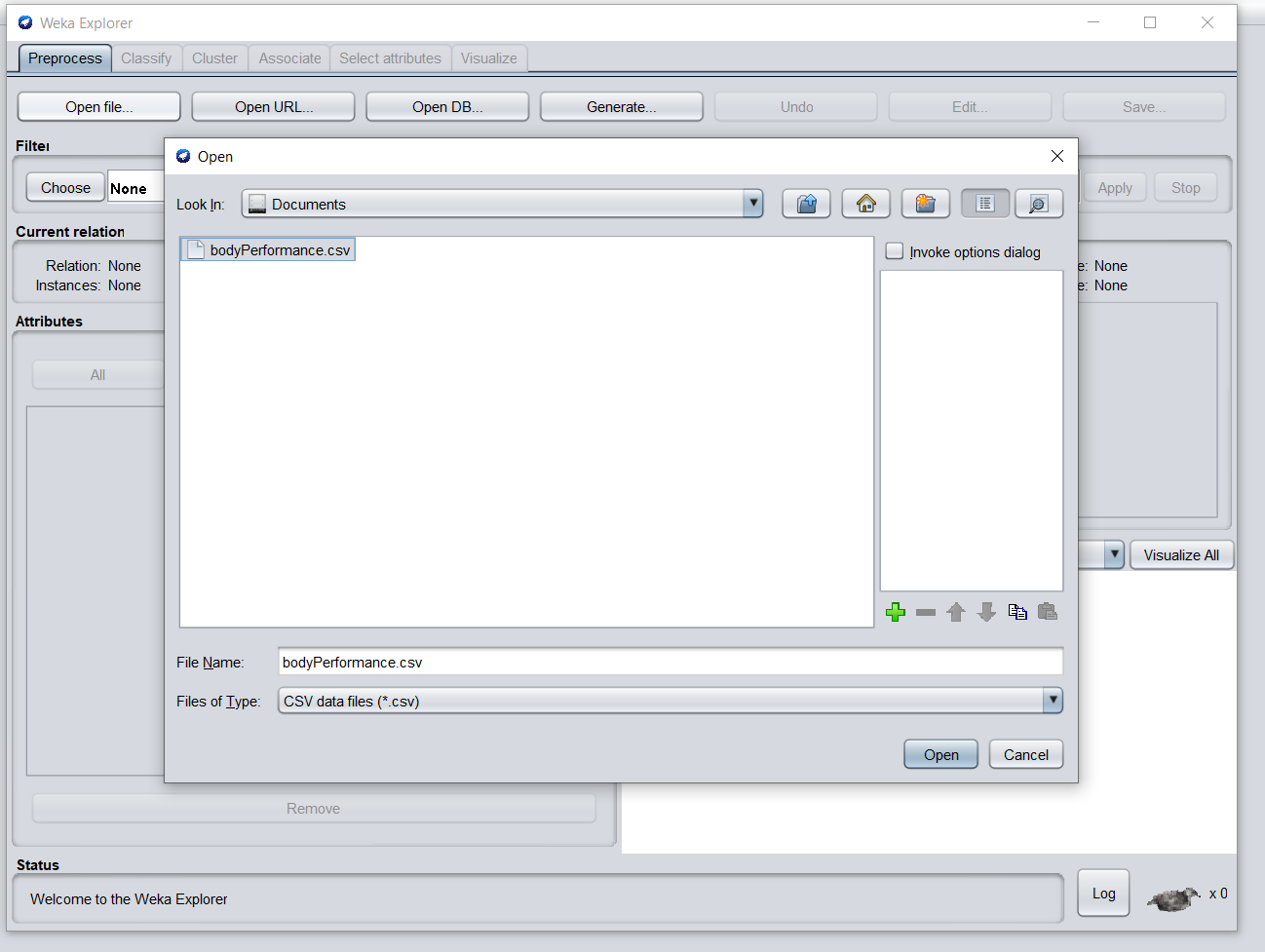
Those are

* Bayes - ( NaiveBayes )
* Lazy - (IBK)
* Trees - (J48)

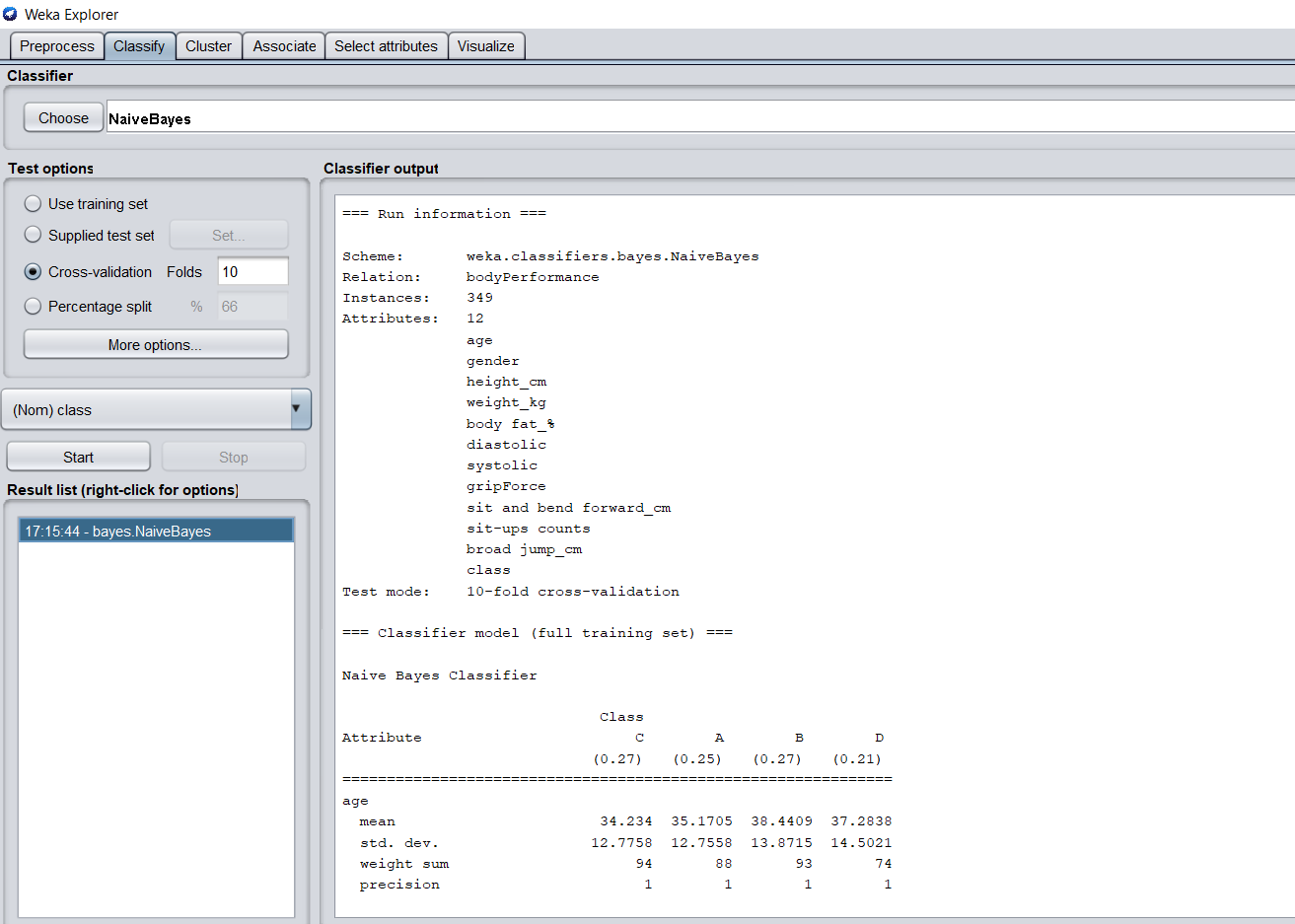
**1.Bayes - ( NaïveBayes )**

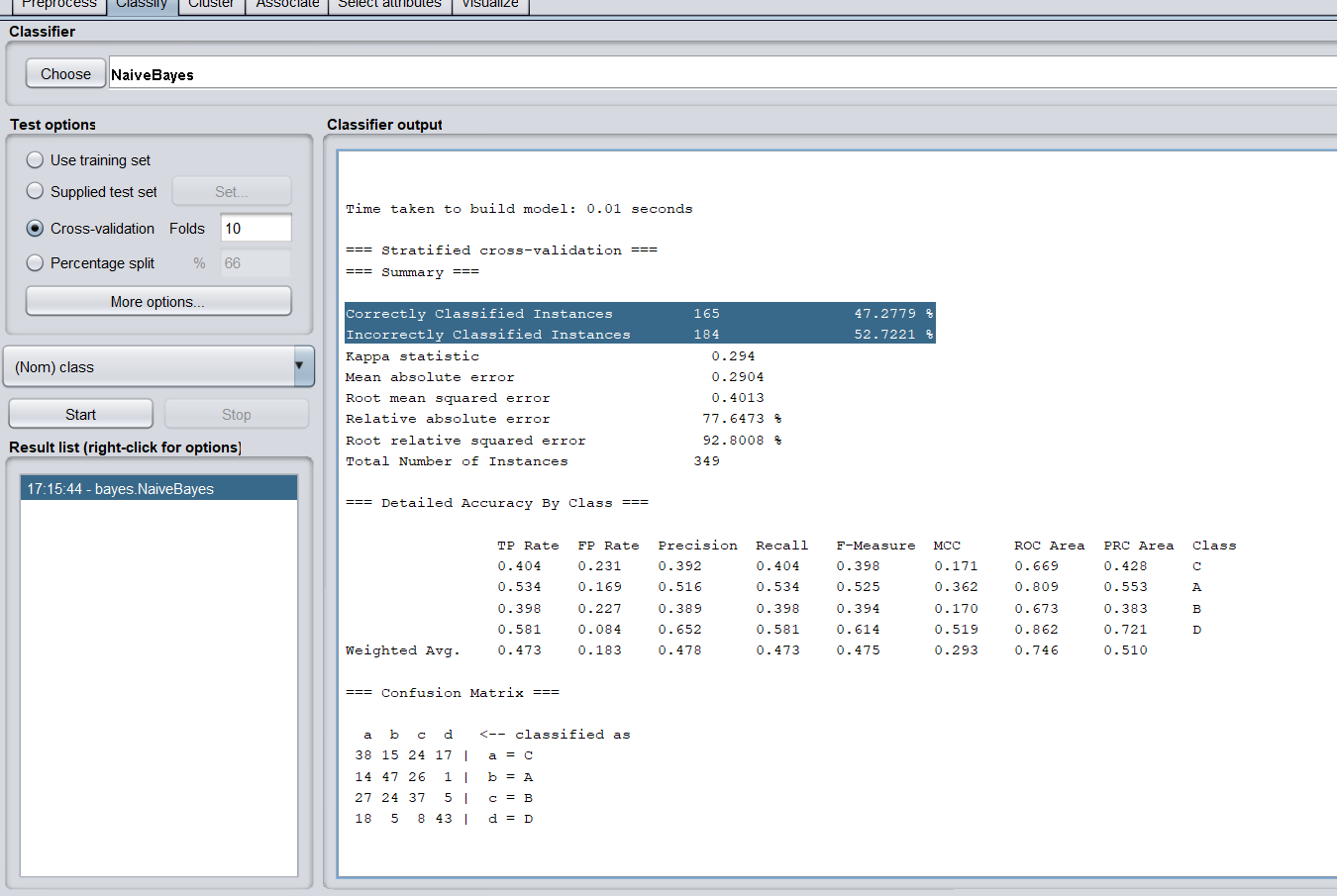
The Bayes theorem is a mathematical method for calculating the likelihood of conditional acceptance. Naive Bayes is a straightforward method for building classifiers. Models that use a finite number of class labels to give class labels to issue situations as vectors of feature values. All naive Bayes classifiers assume that the value of one feature is independent of the value of any other feature, given the class variable.

Step-1: Selection of the csv dataset



Step-2: Selection of Naïve Bayes from the Bayes classifiers section.





**Result**

Time taken to build model: 0.01 seconds

Total Number of Instances 349

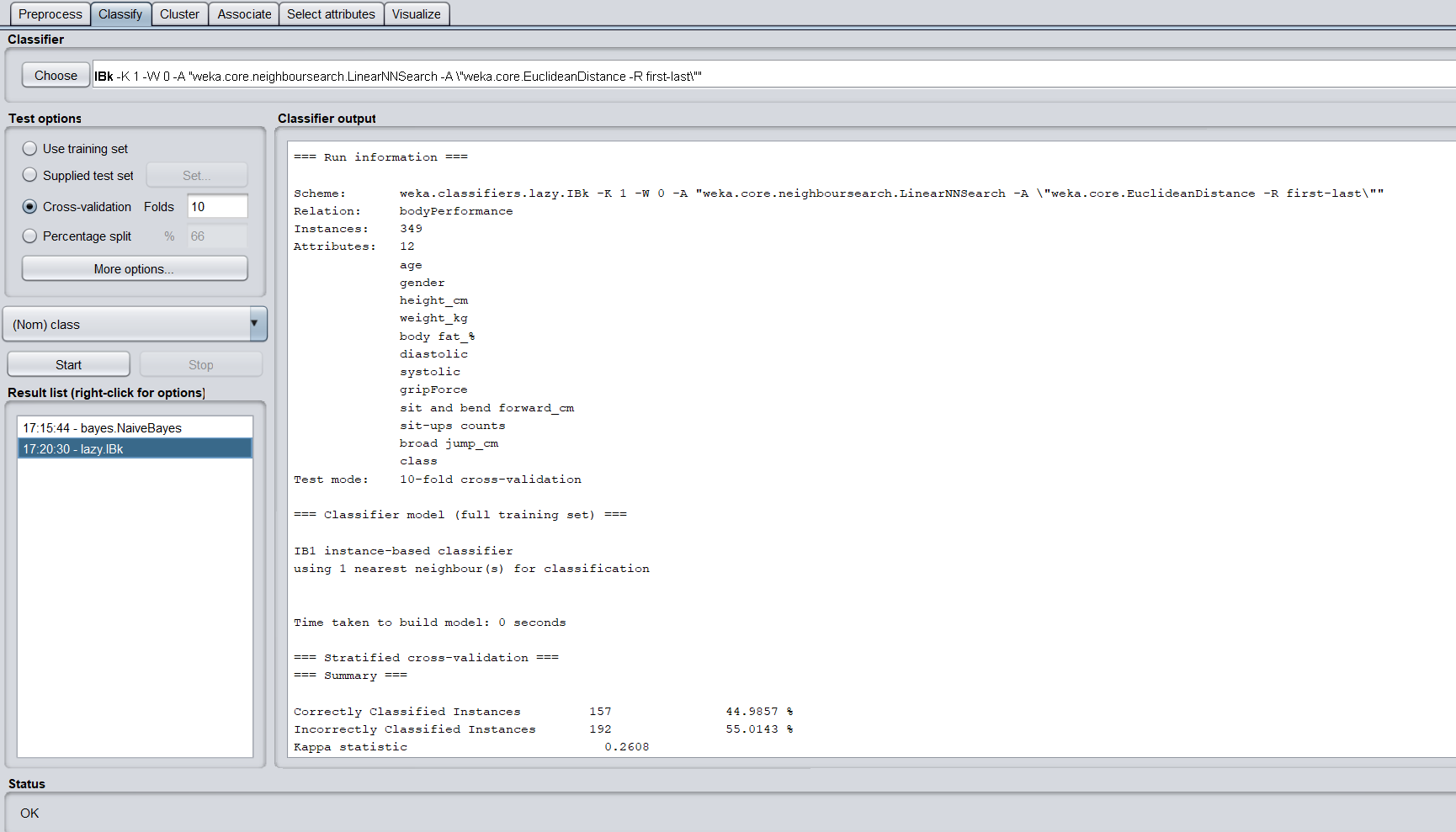
\*Correctly Classified Instances 165 47.2779 %

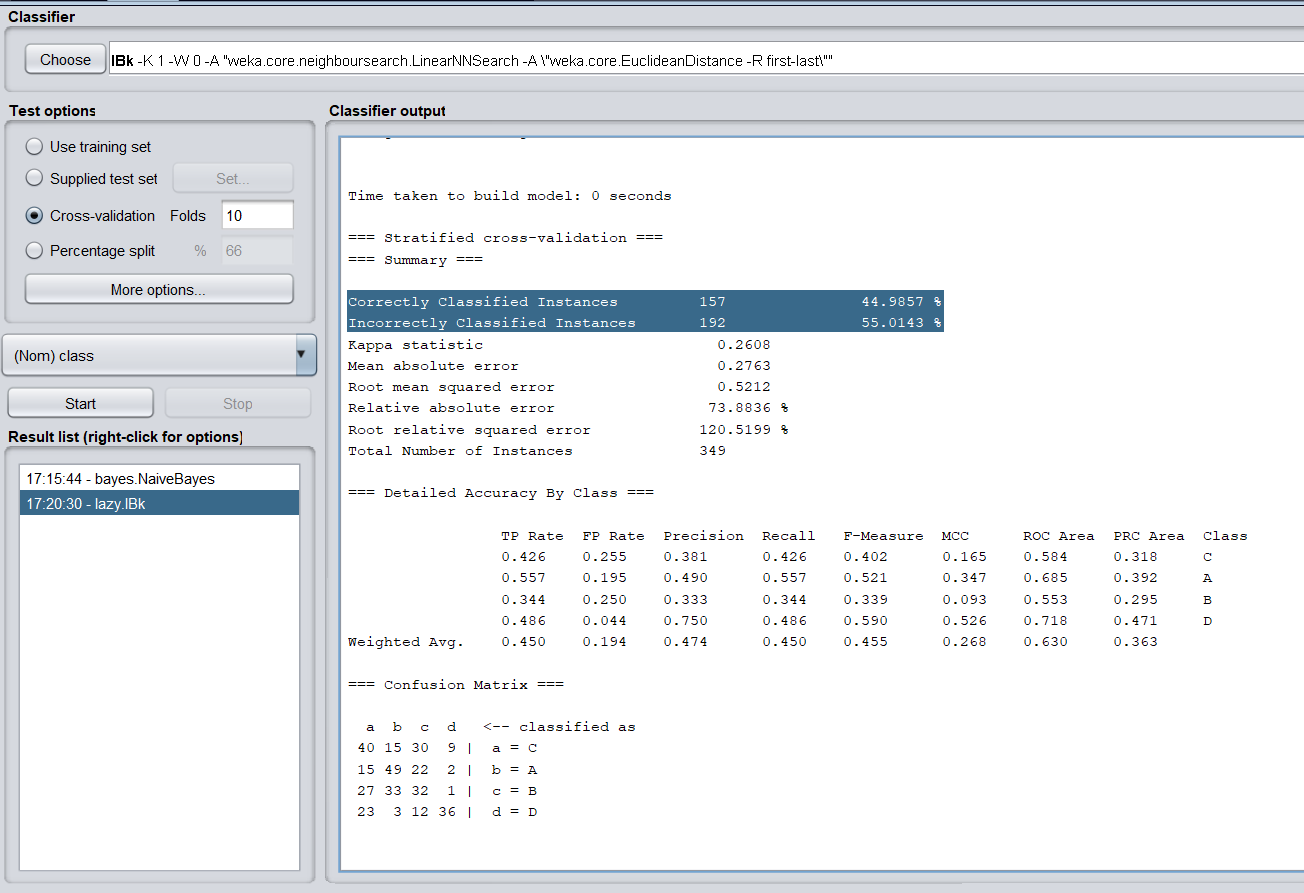
\*Incorrectly Classified Instances 184 52.7221 %

**2.Lazy - (IBK)**

Euclidean distance is a widely used distance measure for continuous variables. Another measure can be used for discrete variables, such as text categorization. For classification and regression, the k-nearest neighbor technique is utilized. In both circumstances, the input is a data set with the k closest training samples. Whether k-NN is used for classification or regression determines the outcome.

Step: Selection of IBK from the Lazy classifiers section





**Result**

Time taken to build model: 0 seconds

Total Number of Instances 349

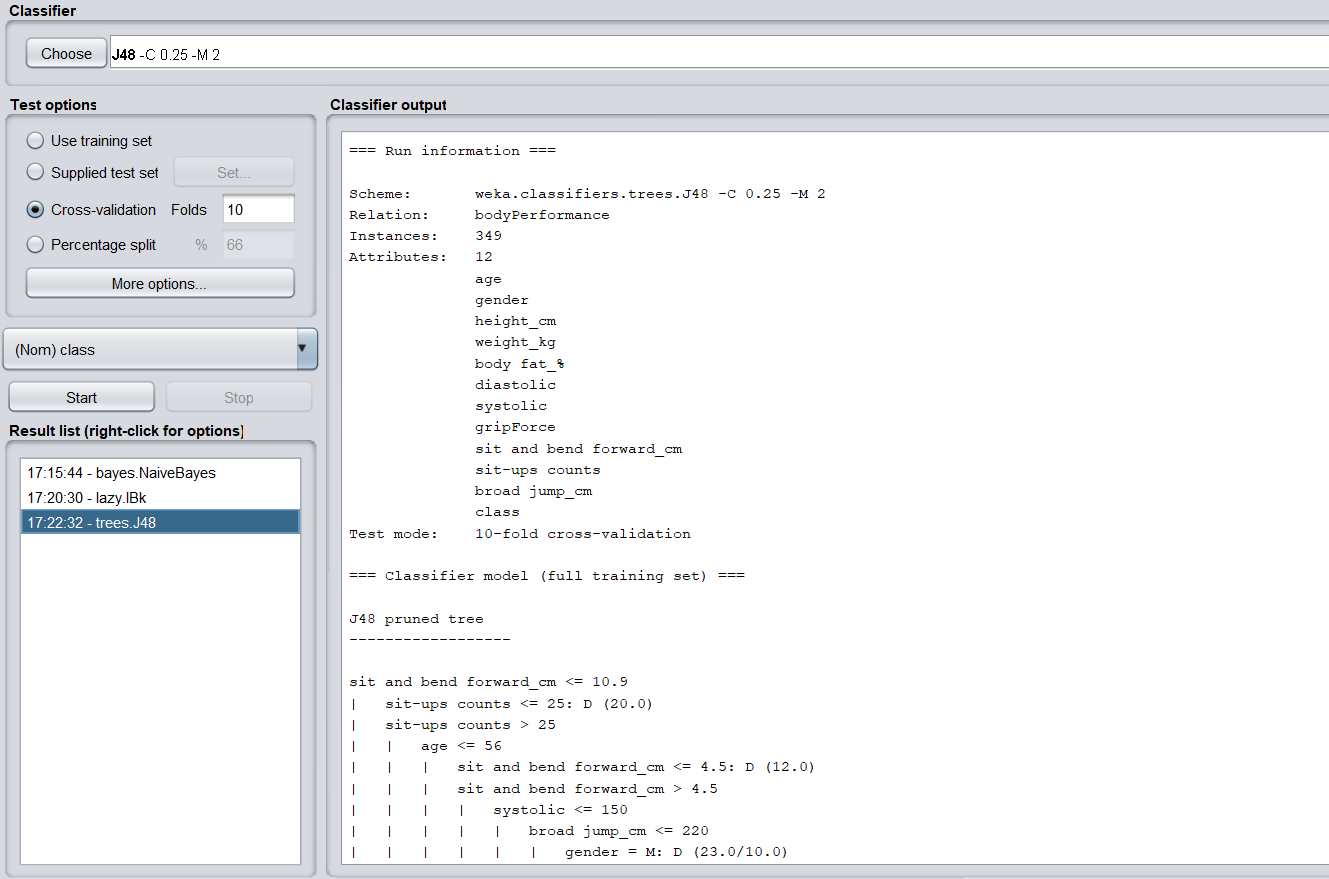
\*Correctly Classified Instances 157 44.9857 %

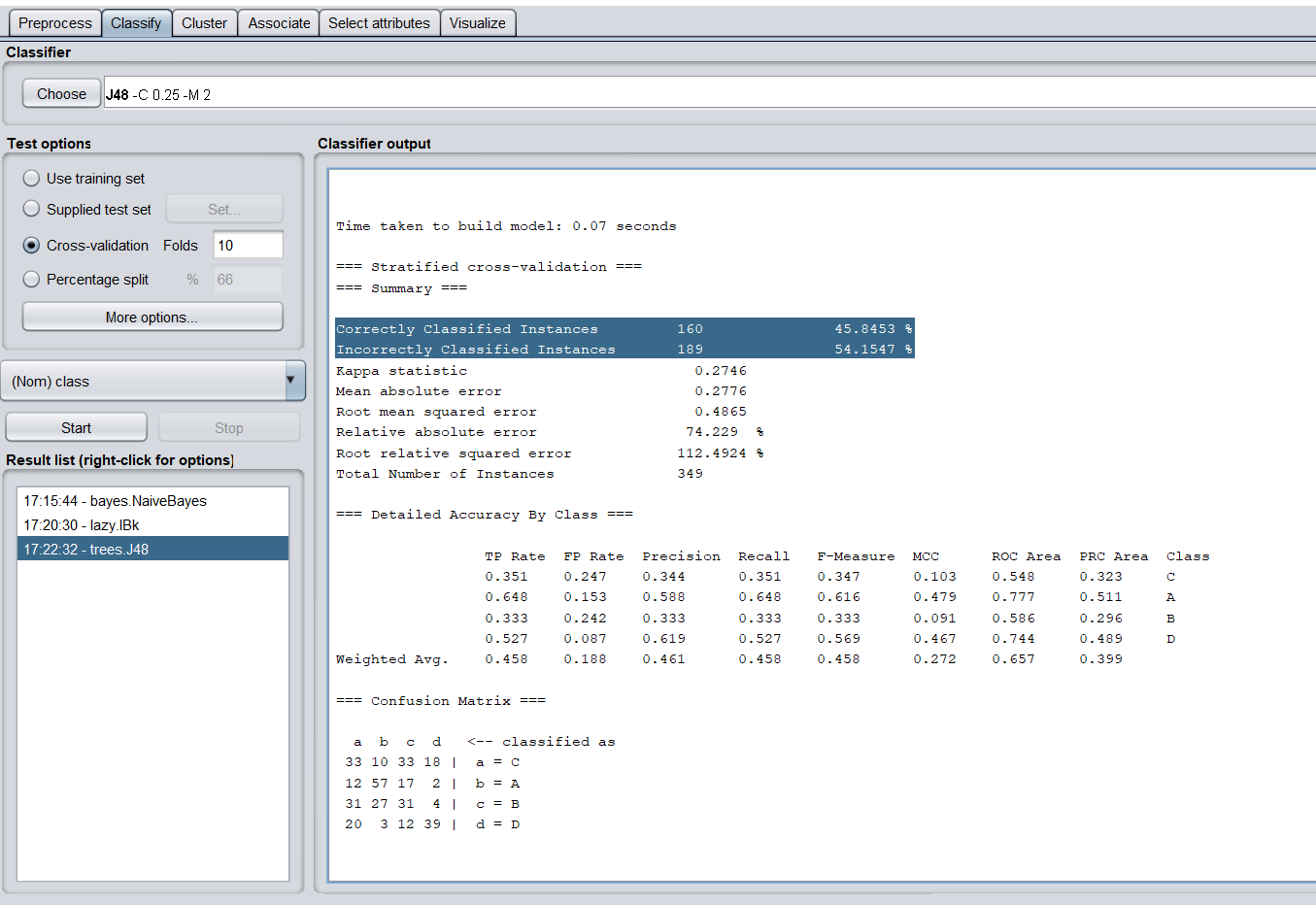
\*Incorrectly Classified Instances 192 55.0143

**3.Trees - (J48)**

A decision tree is a type of tree structure that resembles a flowchart.

Step: Selection of J48 from the Trees classifiers section.





**Result**

Time taken to build model: 0.07 seconds.

Total Number of Instances 349

\*Correctly Classified Instances 160 45.8453 %

\*Incorrectly Classified Instances 189 54.1547 %

**Final Result**

After using Naive Bayes classifiers, the result reveals that 165 instances were successfully identified, with a rate of 47.2779 percent. And with 184 instances erroneously categorized, the percentage is 52.7221 percent. IBK classifiers properly identified 157 instances, with a 44.9857 percent accuracy rate. And with 192 instances erroneously categorized, the percentage is 55.0143 percent. In J48 trees, 160 cases were successfully identified, with a 45.8453 percent success rate. Also, 189 is wrong. The rate of classified incidents is 54.1547 percent**.**

As a result, the accuracy of the Naive Bayes classifiers is higher. The number of right cases is significantly higher than the other two. So Naive Bayes is the best classifier.

**Decision:** Best Classifier – Naïve Bayes.

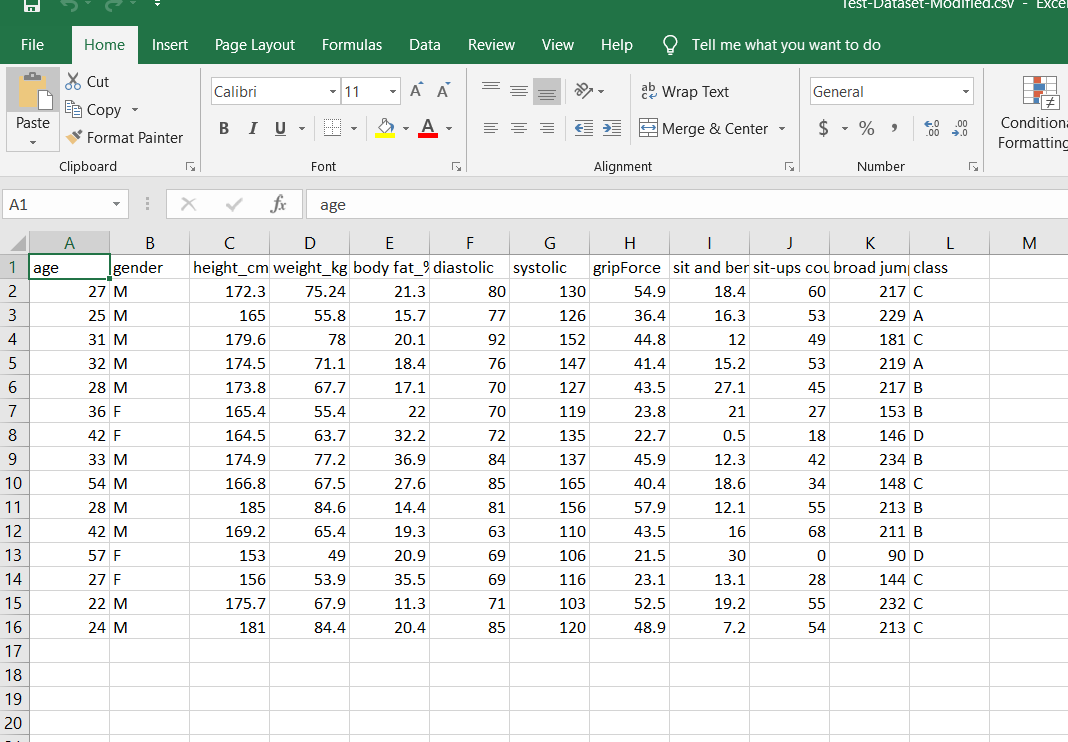
**Discussion**

The best classifier is picked based on accuracy and the percentage of accurate cases. There are no technical concerns while using the classifiers in Weka.

**Task-2**

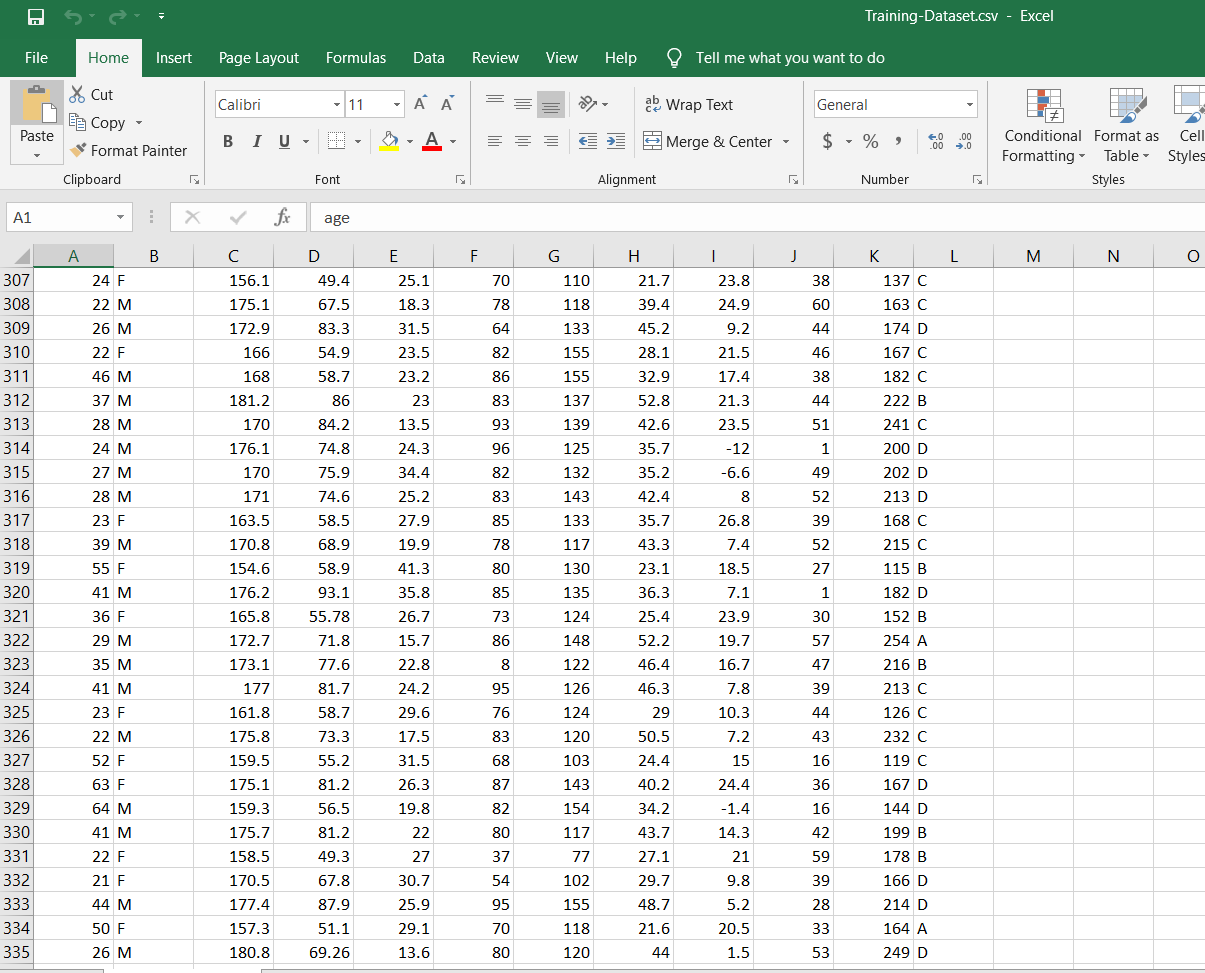
**Introduction**

In task 2, I put our chosen dataset that we utilized in task 1 to the test. I need to prepare two datasets for testing. The first is a test dataset, while the second is a training dataset. I clipped my dataset and picked the top 15 rows as testing data to generate the test dataset. Also, certain numbers were altered to ensure that the testing was accurate.

****

The test dataset is attached as a screenshot.

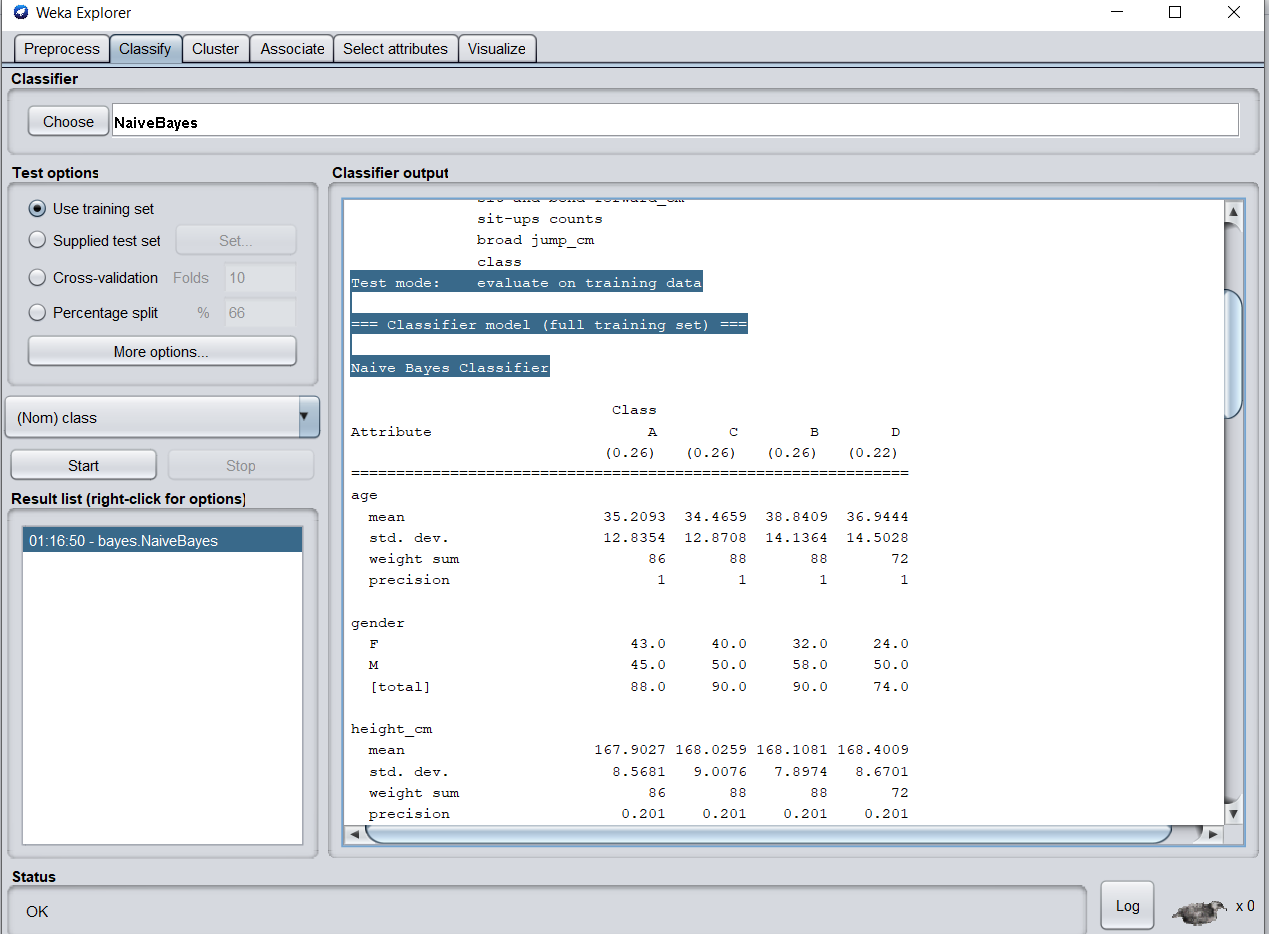
Following the organization of the test dataset, the training dataset is created. Following the clipping of the test dataset, the remaining data is assigned to the training dataset. This training dataset will be used to compare accuracy.

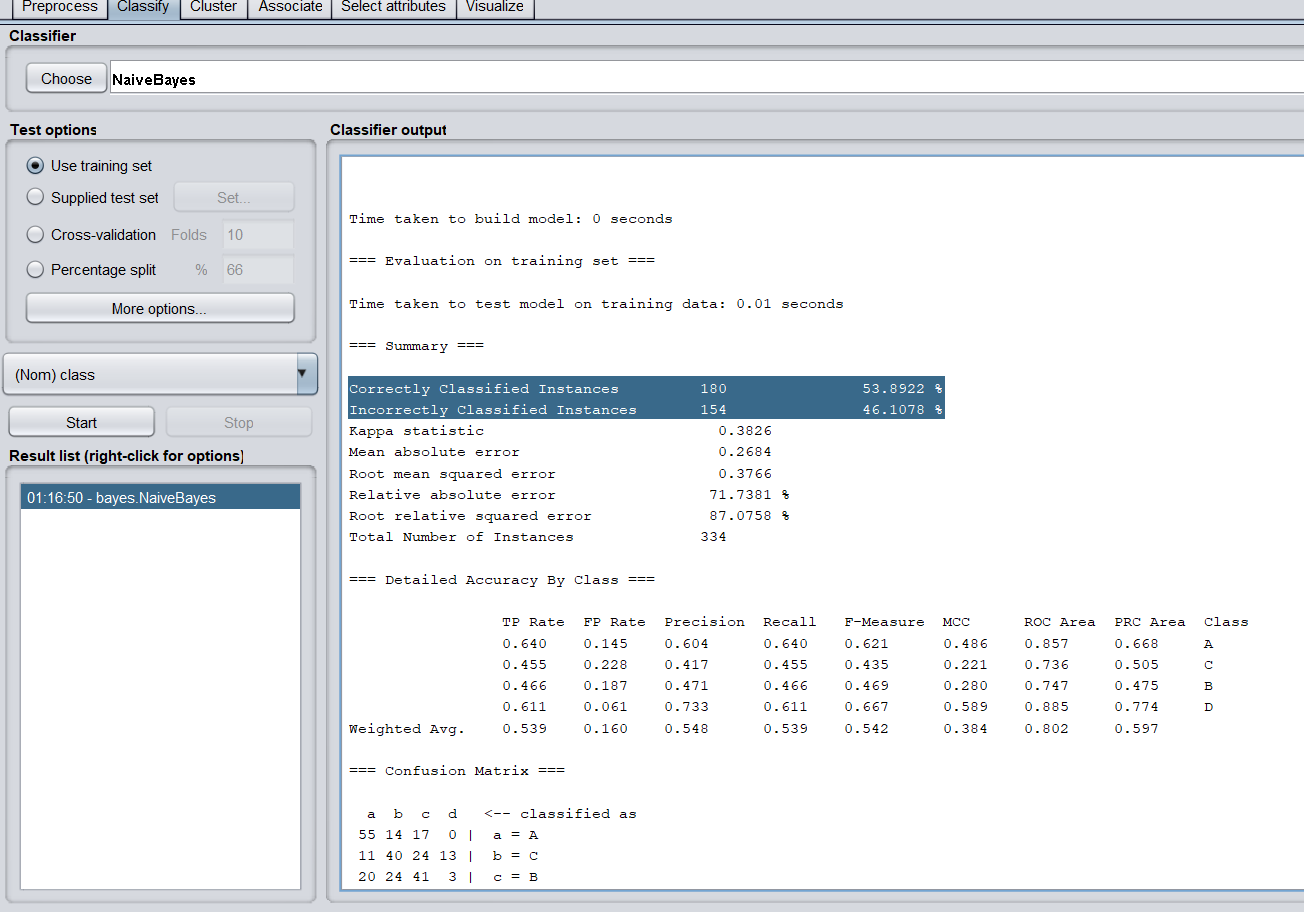


The training dataset is attached as a screenshot.

**Procedure**

Working with the training dataset is the first stage. Following the selection of the training dataset, the best classifier should be used. In task 1, the best classifier for me is Naive Bayes. In my training dataset, I've now used Naive Bayes.



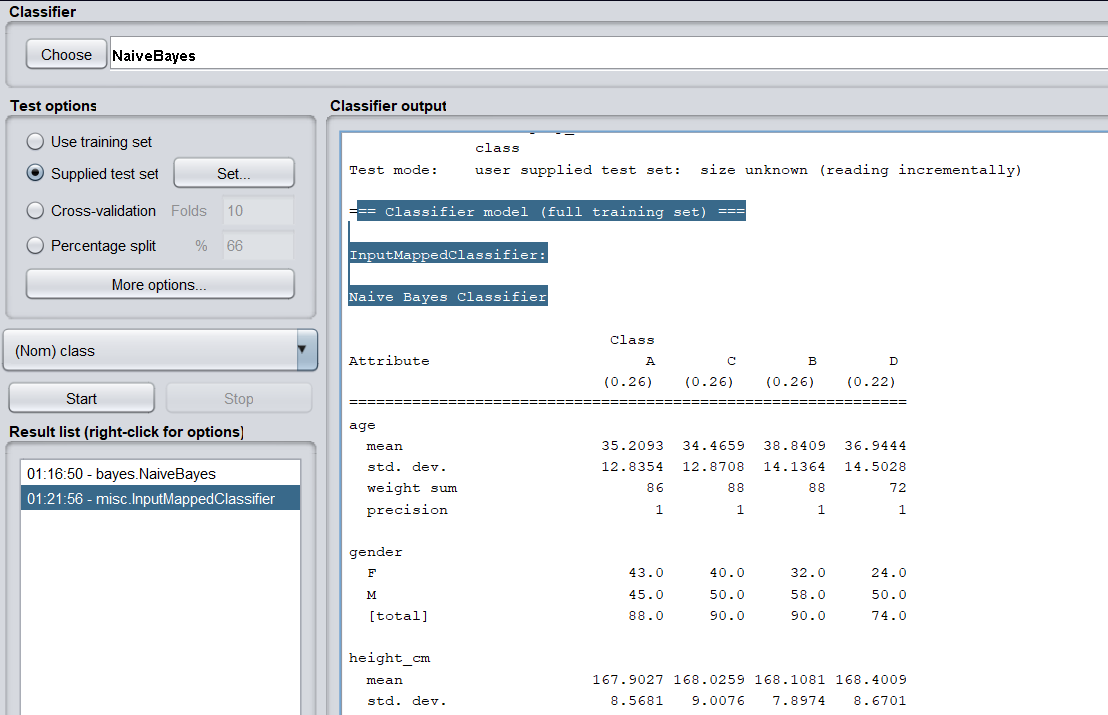


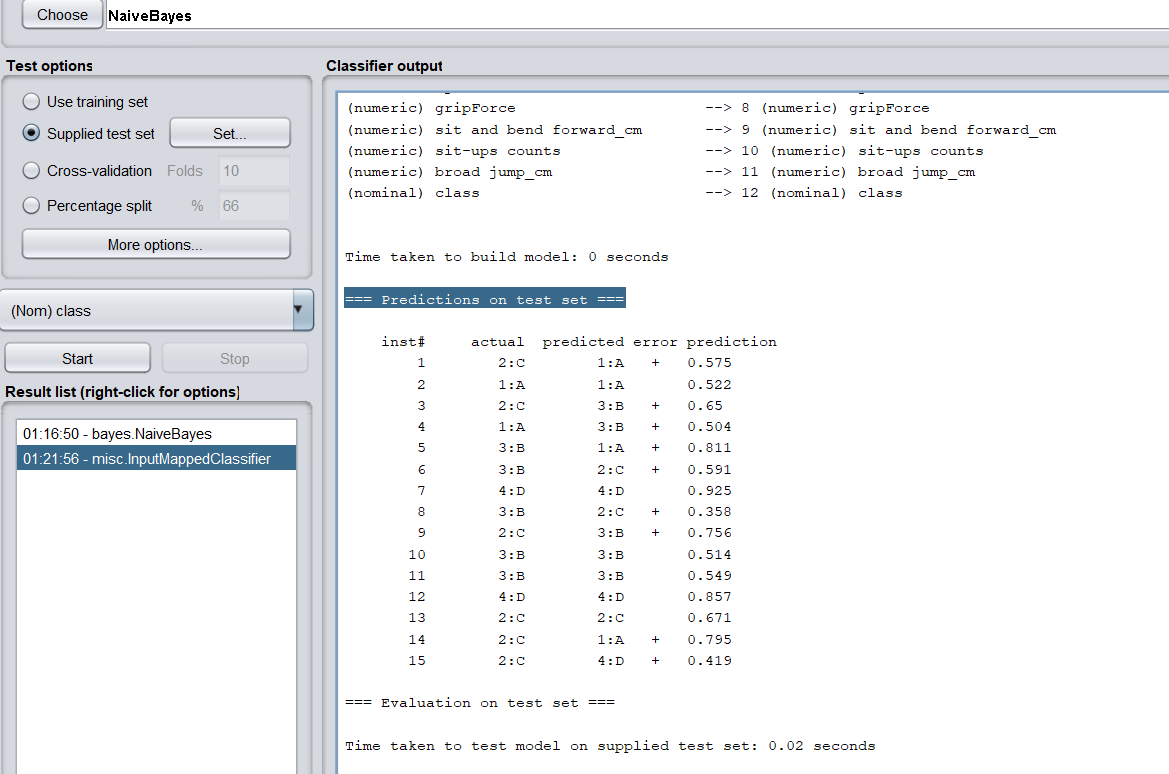
**Result**

Correctly classified instances 180 53.8922%

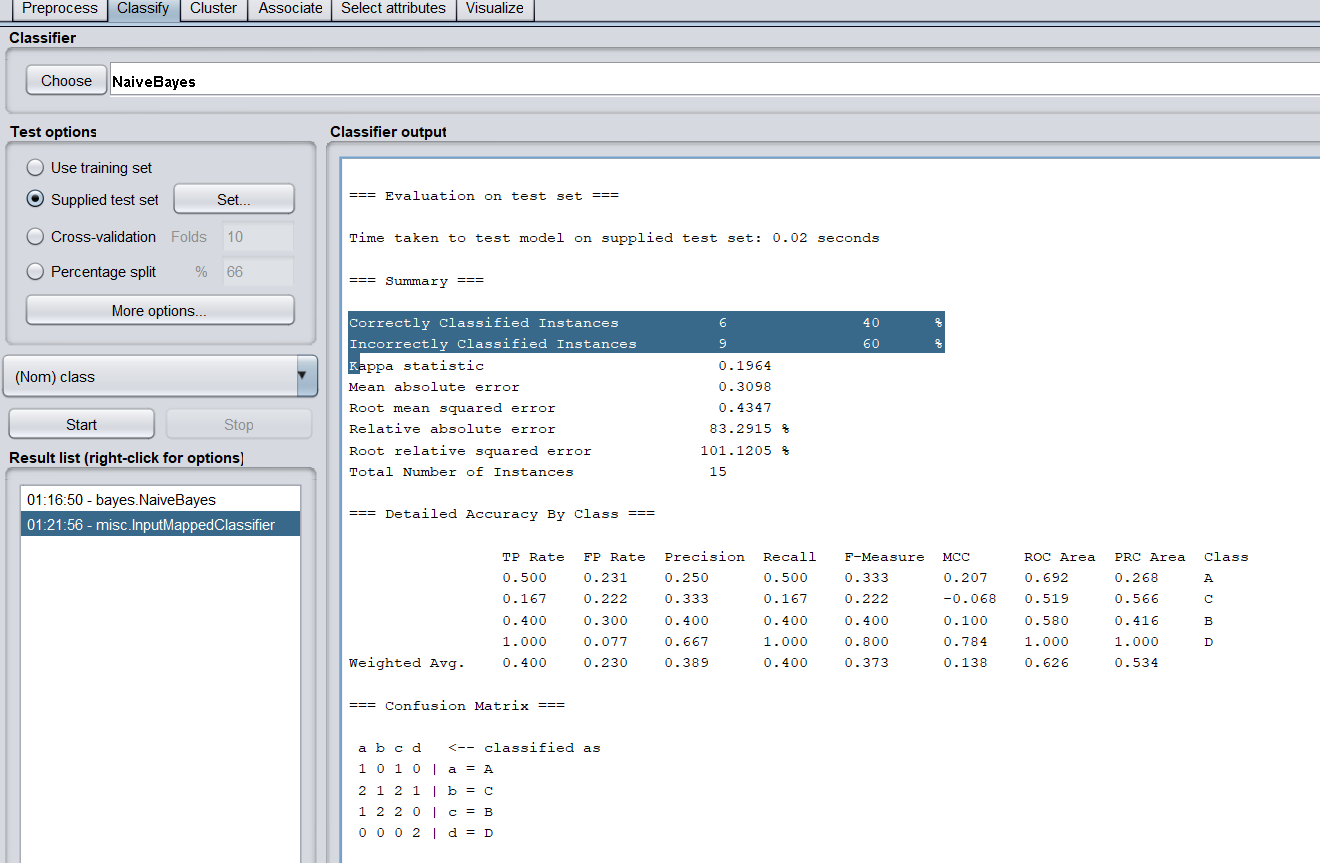
Incorrectly classified instances 154 46.1078%

The accuracy values are kept track of. The following step is to use the testing dataset.





The results of the tests are shown in the attachment. I said that I altered several variables in the test dataset. Those alterations are caught in this word. As well as displaying faults as output. The testing method was completed successfully, and the result was compared to the actual dataset and the predicted dataset.



**Final Result**

Correctly classified instances 6 40%

Incorrectly classified instances 9 60%

**Discussion**

We can see the comparisons between the measurements of the training and testing datasets. By utilizing the Naive Bayes classifier to construct the technique, we were able to observe the correctness of the tests.

**Reference**

* <https://www.kaggle.com/kukuroo3/body-performance-data>
* Weka 3.8.5