# Ahmad Ali

# ID: 13473 SE-07

# Data Mining

# Submitted To: Sir Zain Shukat

|  |  |
| --- | --- |
| Classification | Regression |
| * The discovery of model or functions where the mapping of objects is done into predefined classes. | * A devised model in which the mapping of objects is done into values. |
| * Discrete values | * Continuous values |
| * Decision tree, logistic regression, etc. | * Regression tree (Random forest), Linear regression, etc. |
| * Unordered | * Ordered |
| * Measuring accuracy | * Measurement of root mean square error |

**Q1:**

**Ans:**

**Key Differences Between Classification and Regression**

1. The Classification process models a function through which the data is predicted in discrete class labels. On the other hand, regression is the process of creating a model which predict continuous quantity.
2. The classification algorithms involve decision tree, logistic regression, etc. In contrast, regression tree (e.g. Random forest) and linear regression are the examples of regression algorithms.
3. Classification predicts unordered data while regression predicts ordered data.
4. Regression can be evaluated using root mean square error. On the contrary, classification is evaluated by measuring accuracy.

Classification example:

Suppose from your past data (train data) you come to know that your best friend likes the above movies. Now one new movie (test data) released. Hopefully, you want to know your best friend like it or not. If you strongly confirmed about the chances of your friend like the move.  You can take your friend to a movie this weekend.

If you clearly observe the problem it is just whether your friend like or not. Finding a solution to this type of problem is called as classification. This is because we are classifying the things to their belongings (yes or no, like or dislike). Keep in mind here we are forecasting target class (classification) and the other thing this classification belongs to supervised learning. This is because you are learning this from your train data.

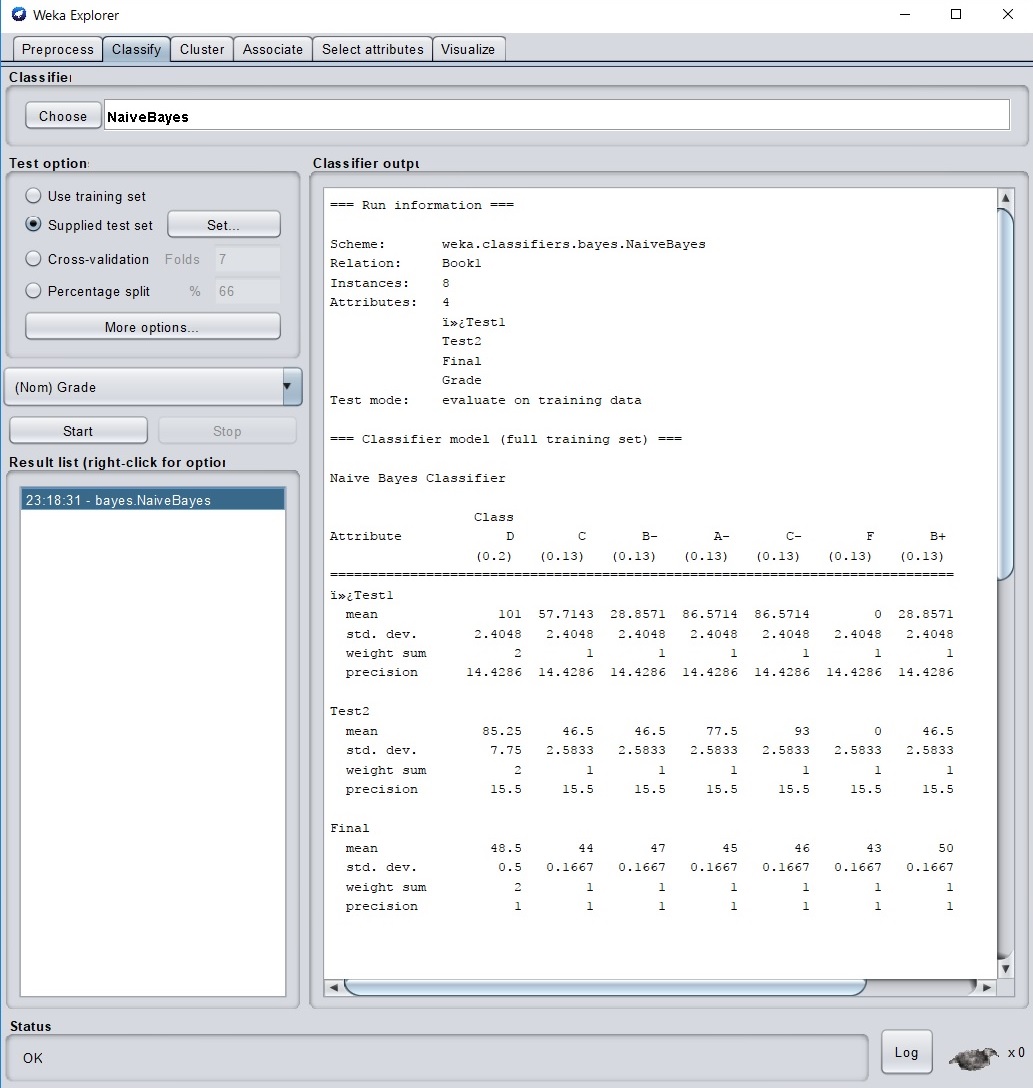
In this case, the problem is a binary classification in which we have to predict whether output belongs to class 1 or class 2 (class 1: yes, class 2: no). As we have discussed earlier, we can use classification for predicting more classes too. Like (color prediction: red,green,blue,yellow,orange)

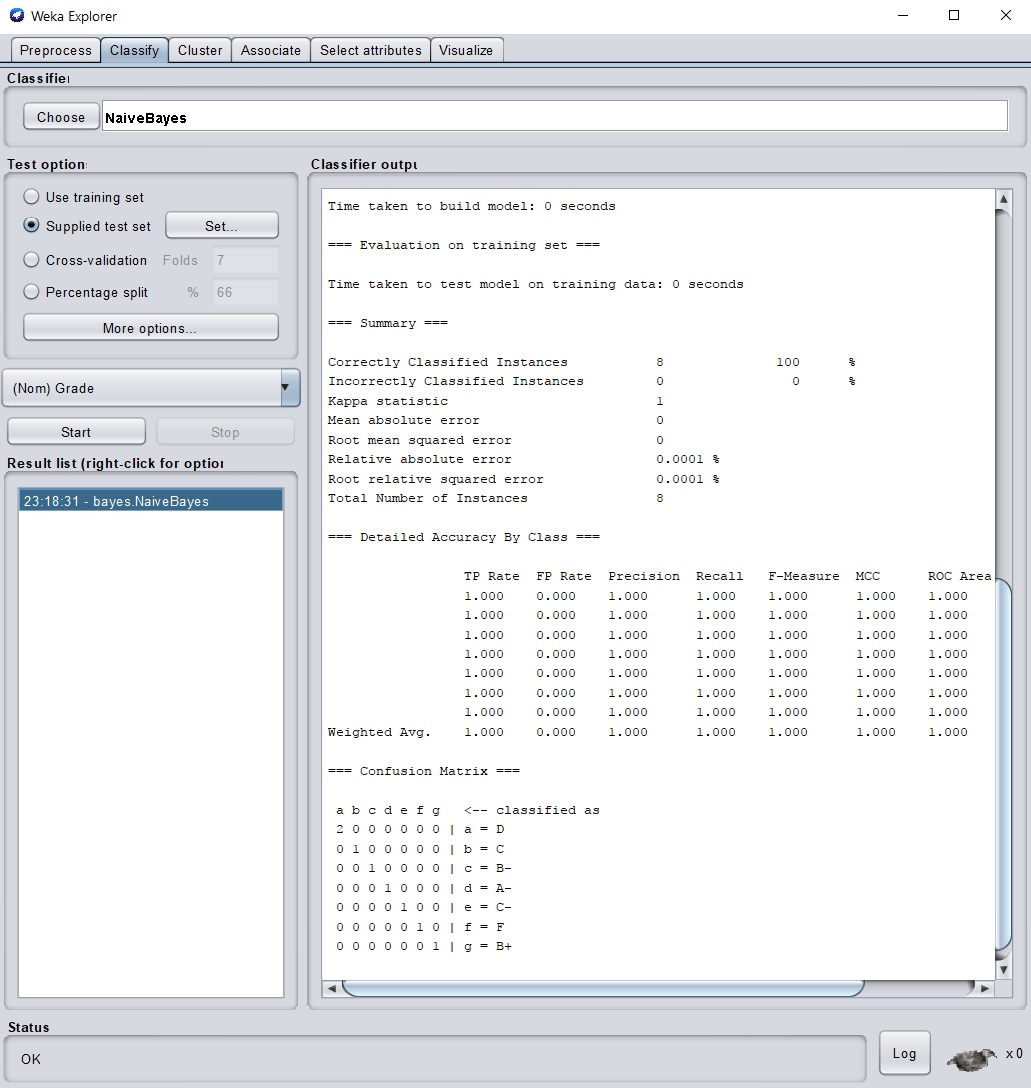
Regression example:

Suppose from your past data (train data) you come to know that your best friend likes the above movies. You also know how many times each particular movie seen by your friend. Now one new movie (test data) released. Now you are going to find how many times this newly released movie will your friend watch. It could be 5 times, 6 times,10 times etc…

If you clearly observe the problem is about finding the count, sometimes we can say this as predicting the value. Keep in mind, here we are forecasting a value (prediction ) and the other thing this prediction also belongs to supervised learning. This is because you are learning this from you train data.

Q2:

Ans:



Q3:

Ans:

* First download any dataset from any source you like.

(make sure it is in .arff extension)

* Open Weka software.
* Click on the explorer button.
* Then click on open file button on top left corner.
* Browse to file you’ve just downloaded
* Add filters if you want.
* Then click on classify button on top.
* Apply various classifier algorithms (naive Bayes, J48 etc.).

