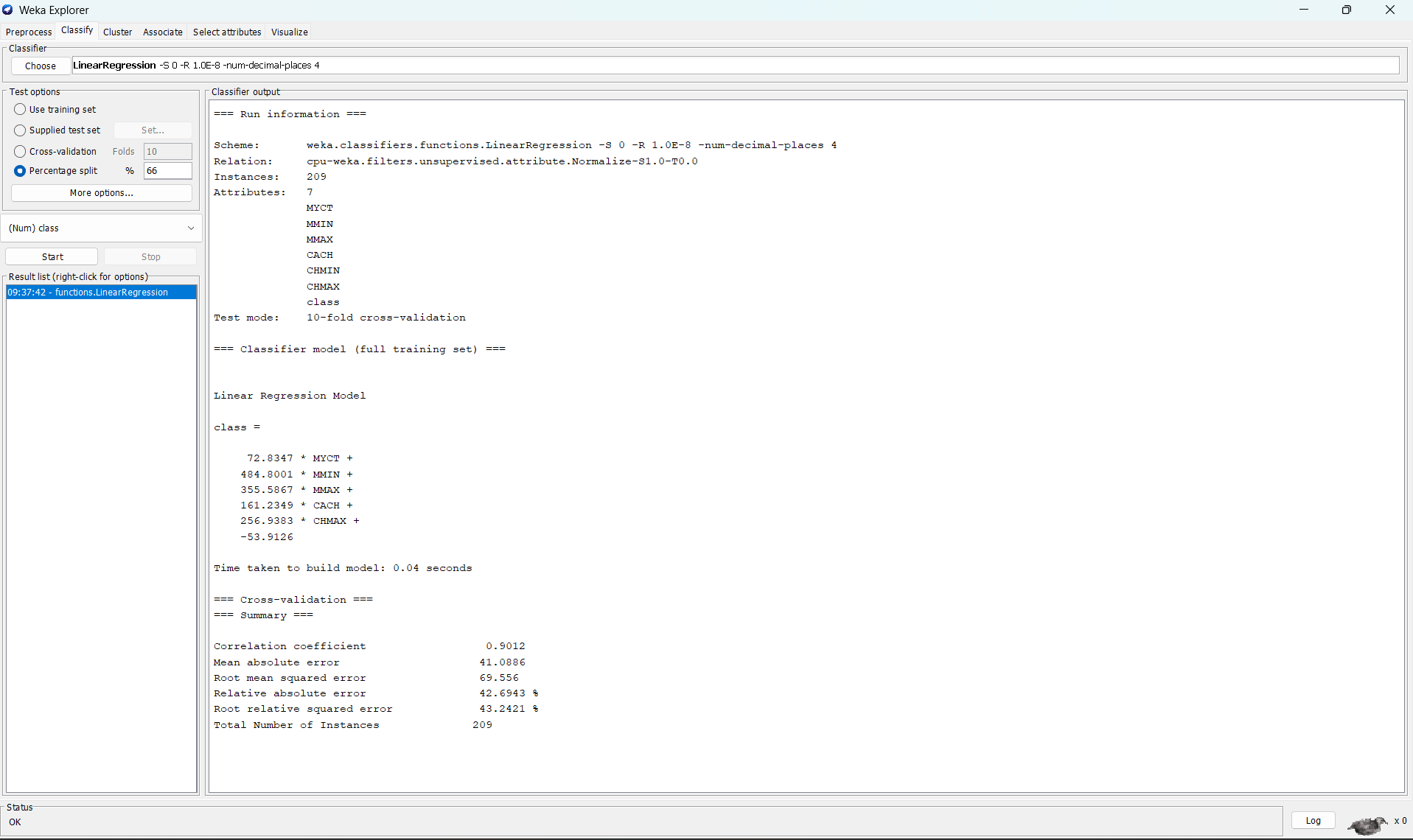
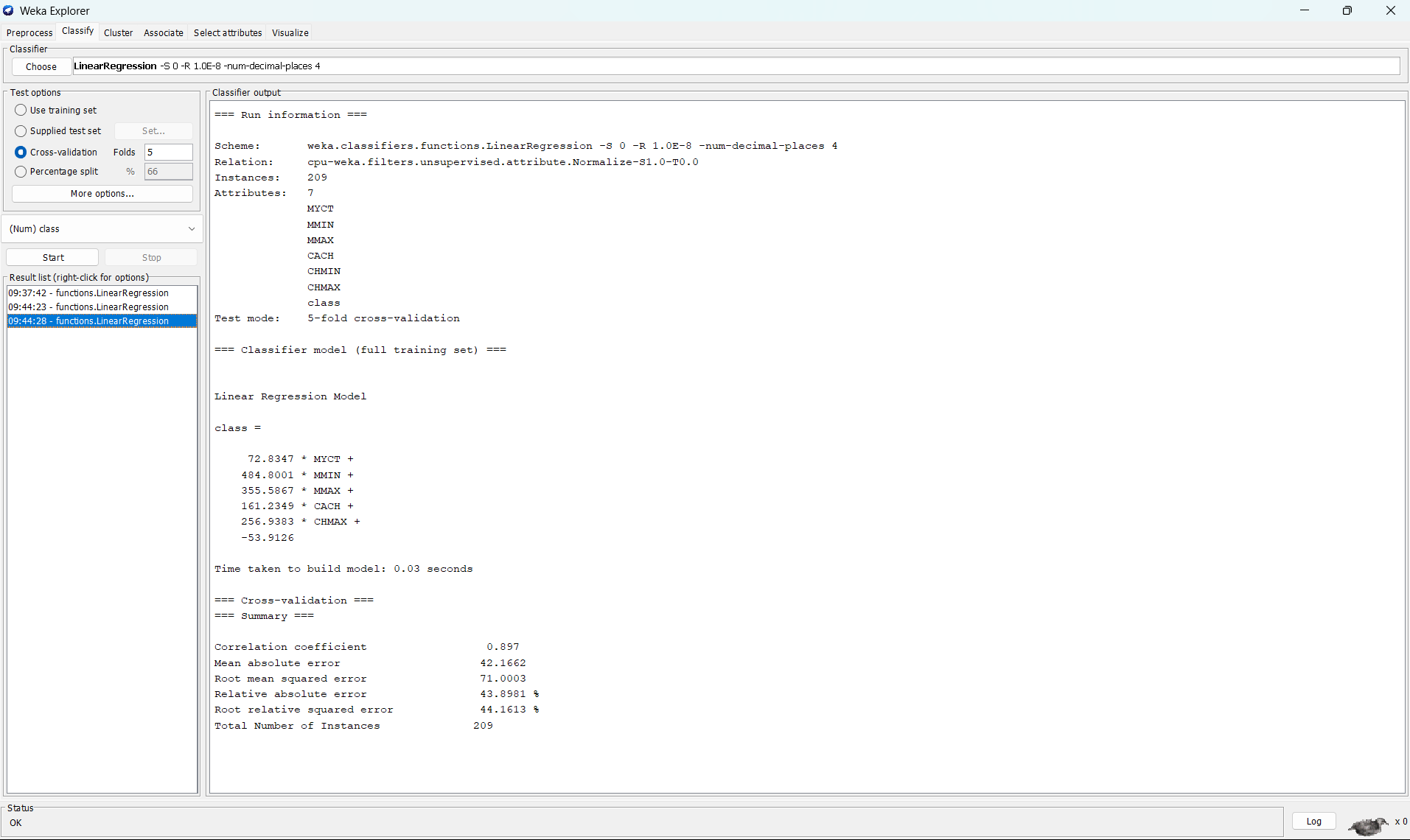
Exp 1

Exploration of WEKA (Waikato Environment for Knowledge Analysis) tool for Regression task.

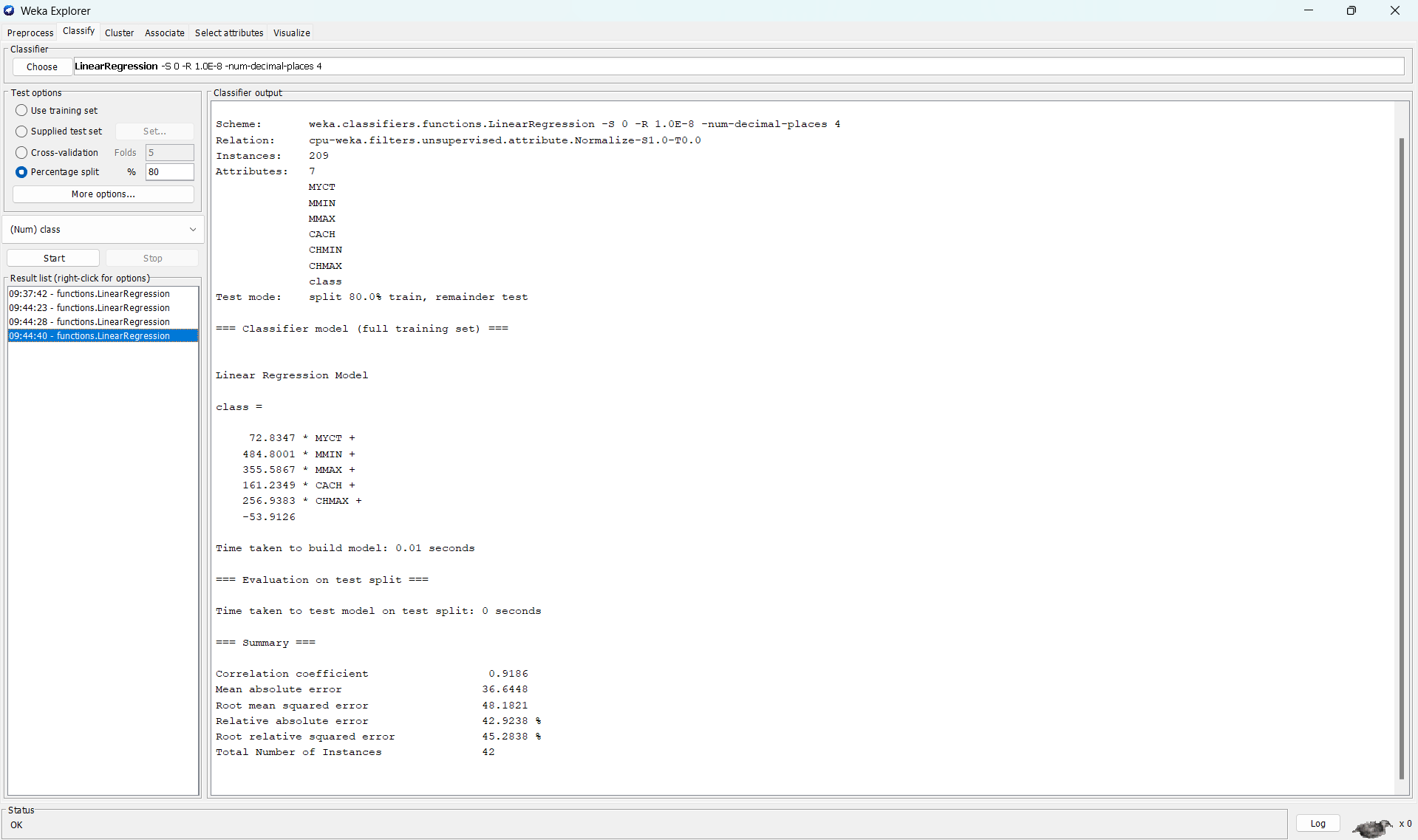
Folds(10)



Folds(5)



split(80%)



**Exp 2**

**Exploration of WEKA tool for Classification task.**

**TP Rate:** rate of true positives (instances correctly classified as a given class)

**FP Rate:** rate of false positives (instances falsely classified as a given class)

**Precision:** proportion of instances that are truly of a class divided by the total instances classified as that class

**Recall:** proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate)

**F-Measure:** A combined measure for precision and recall calculated as 2 \* Precision \* Recall / (Precision + Recall)

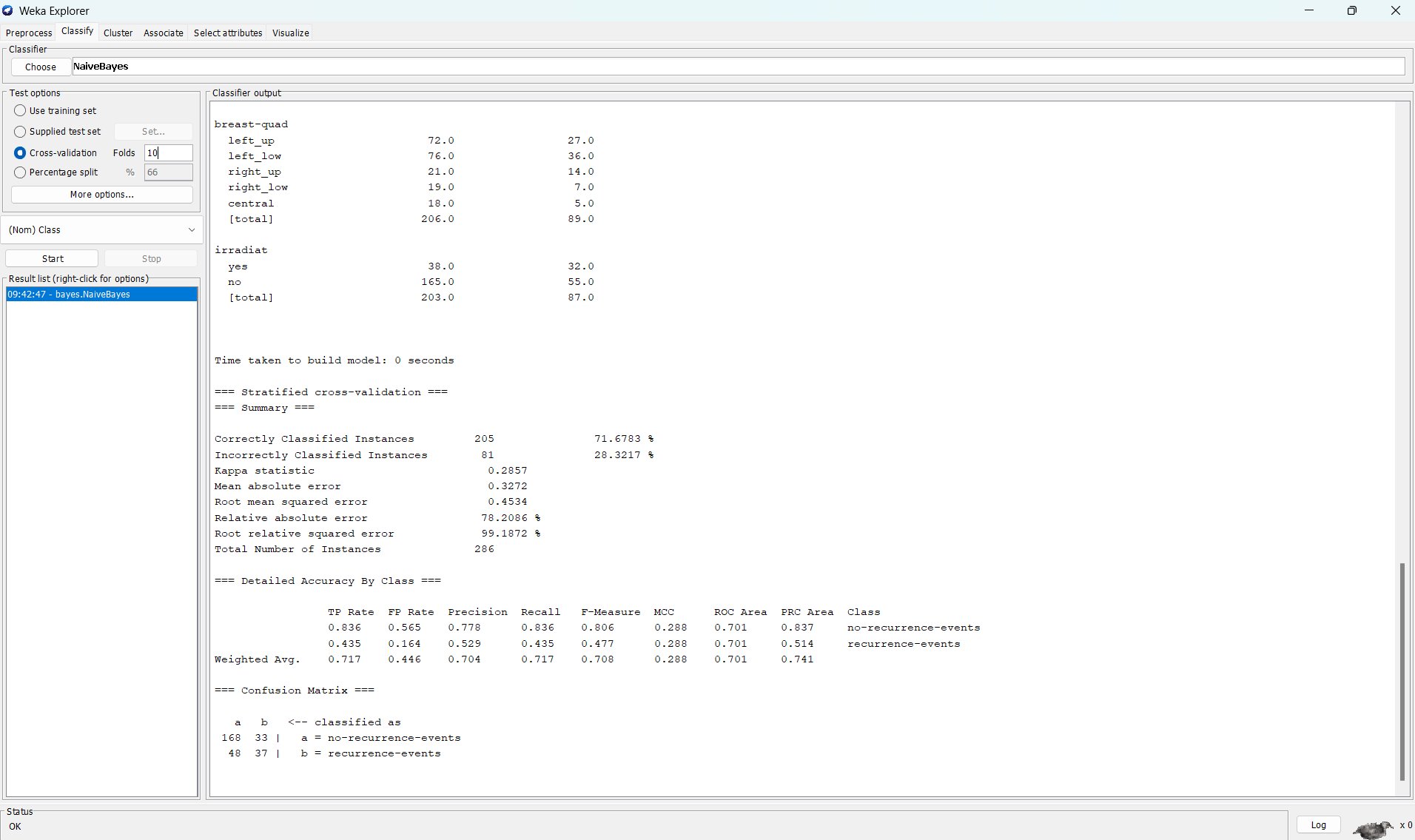
**MCC:** It is used in machine learning as a measure of the quality of binary (two-class) classifications. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes

**ROC:** ( Receiver Operating Characteristics) area measurement: One of the most important values output by Weka. They give you an idea of how the classifiers are performing in general

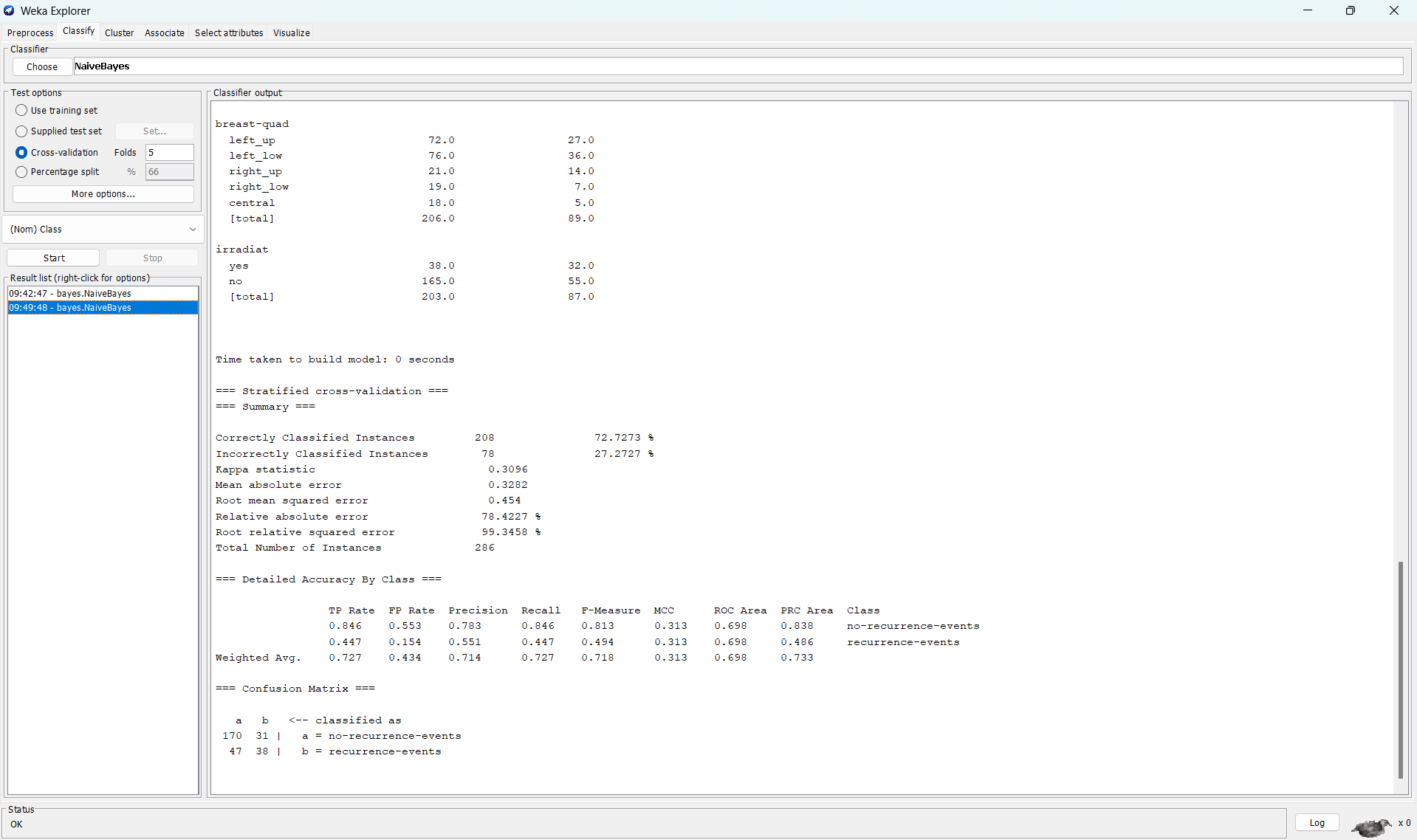
**PRC( Precision Recall) area :** The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets

1. Naive Bayes

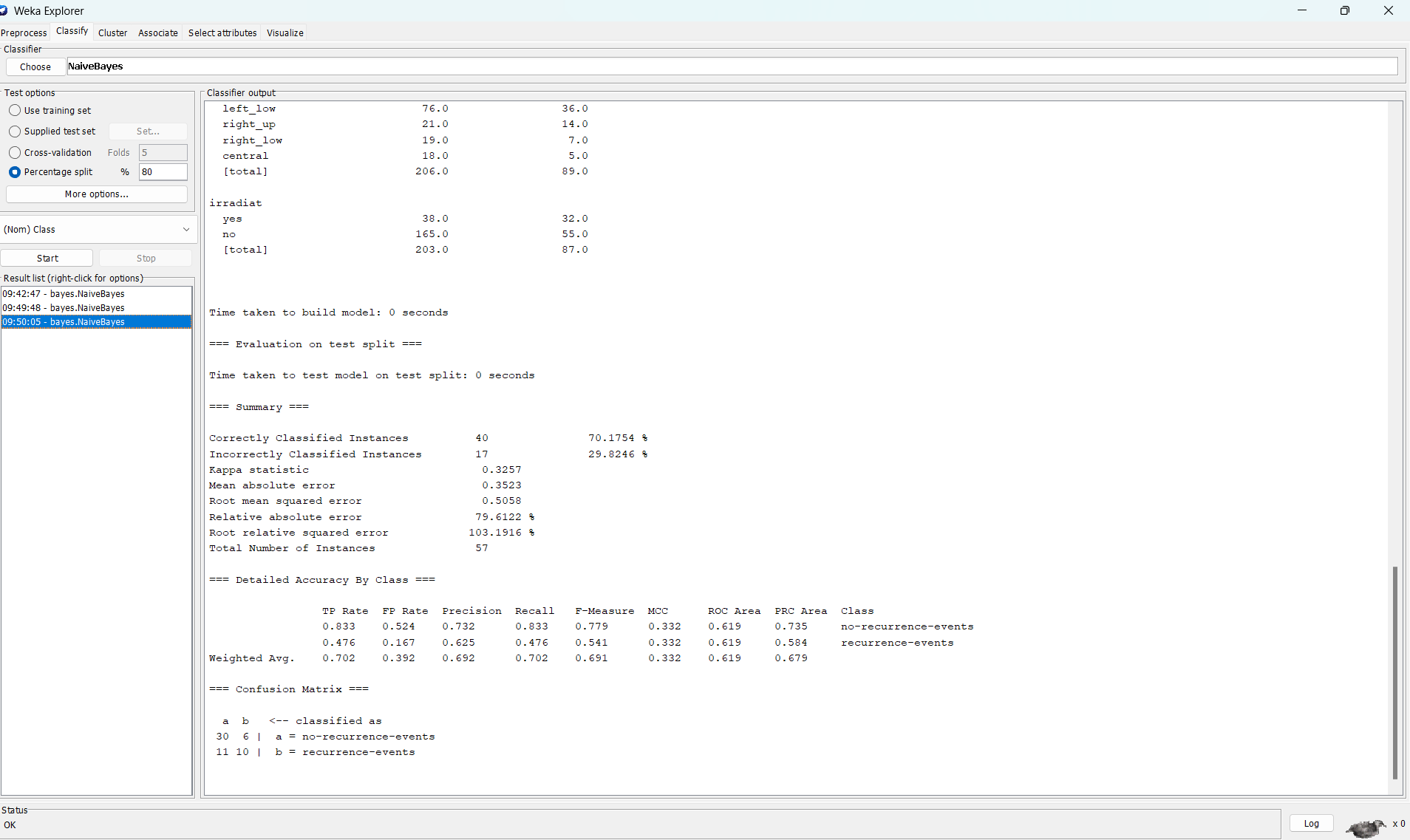
Folds(10)



Folds(5)

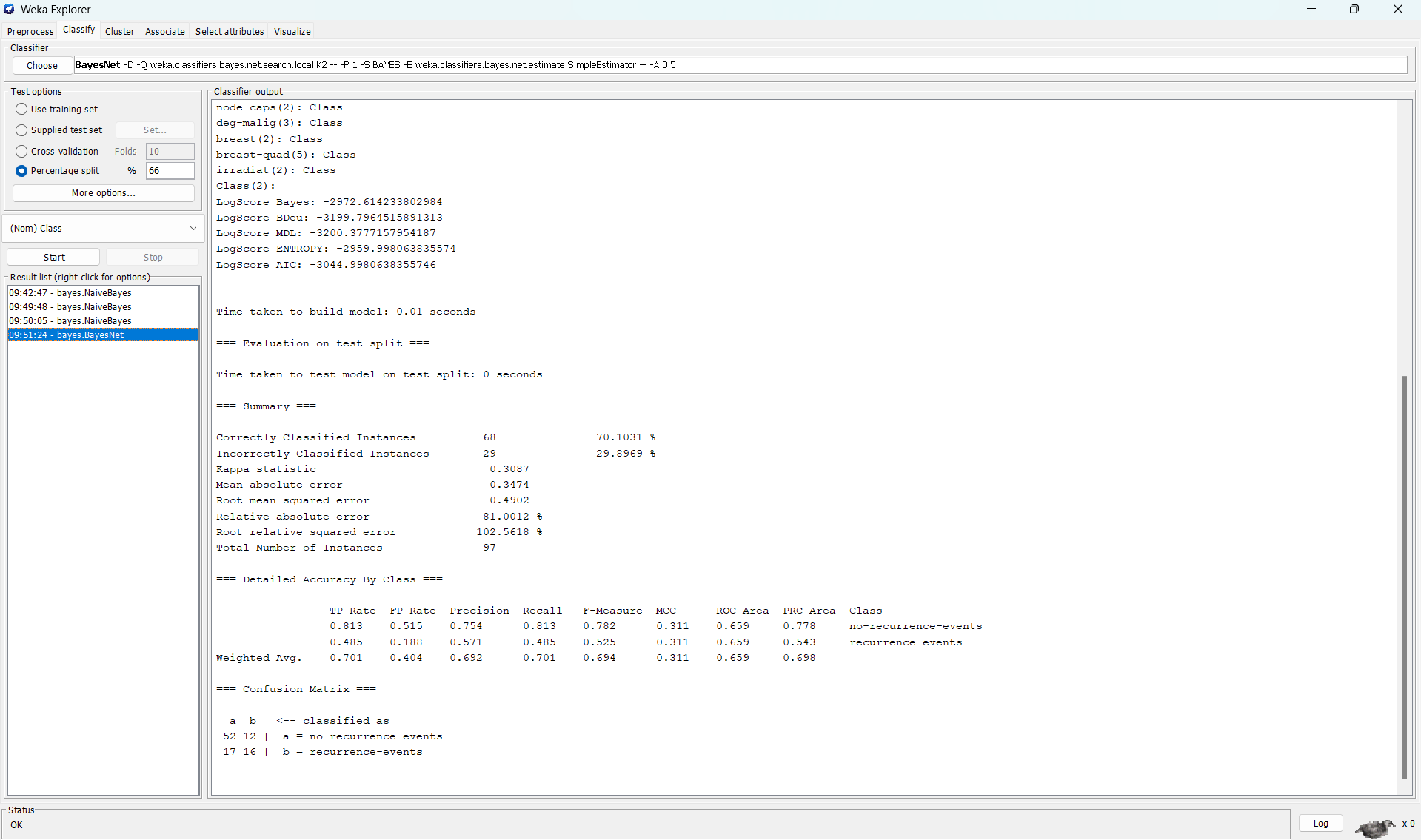


split(80%)

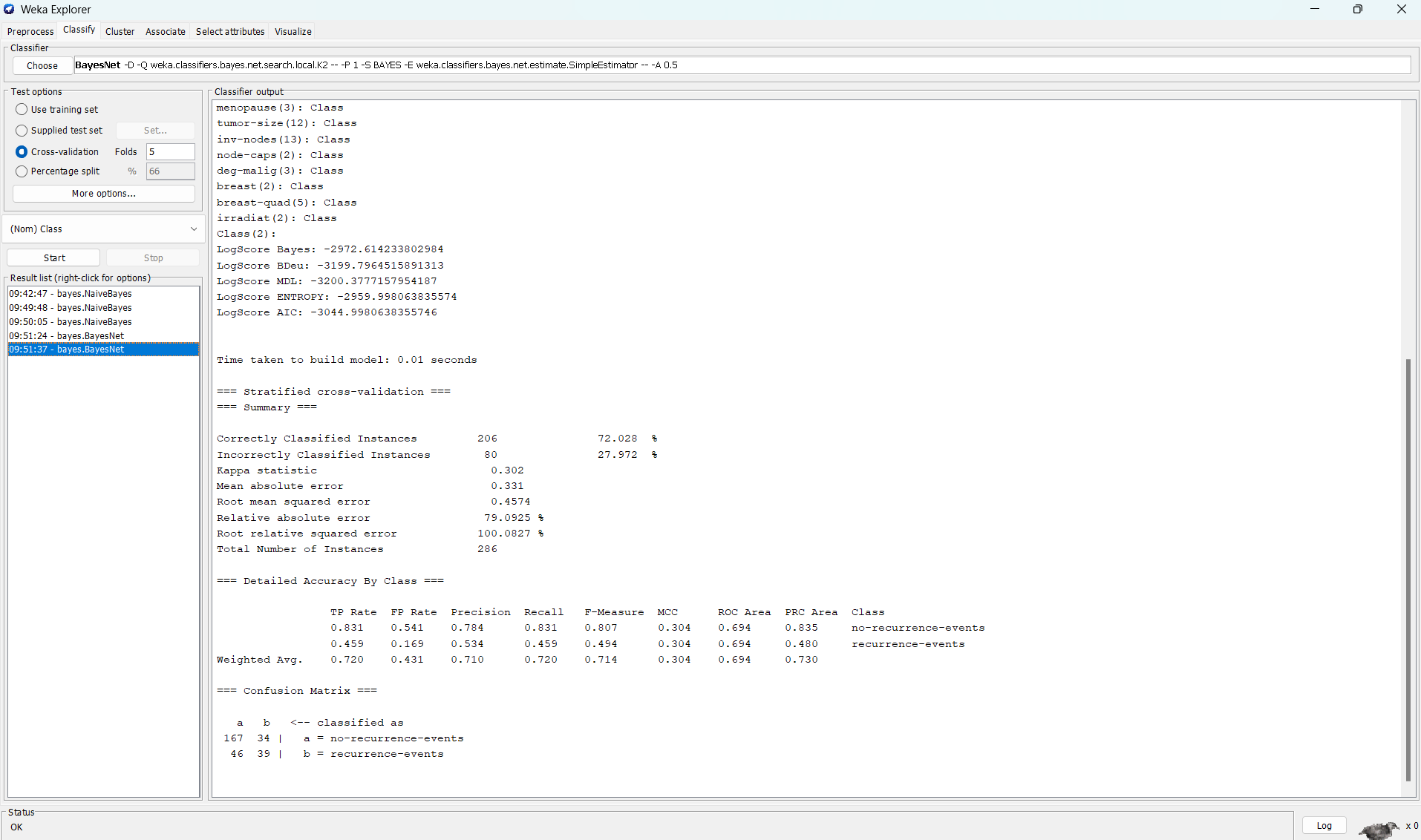


1. Bayes net

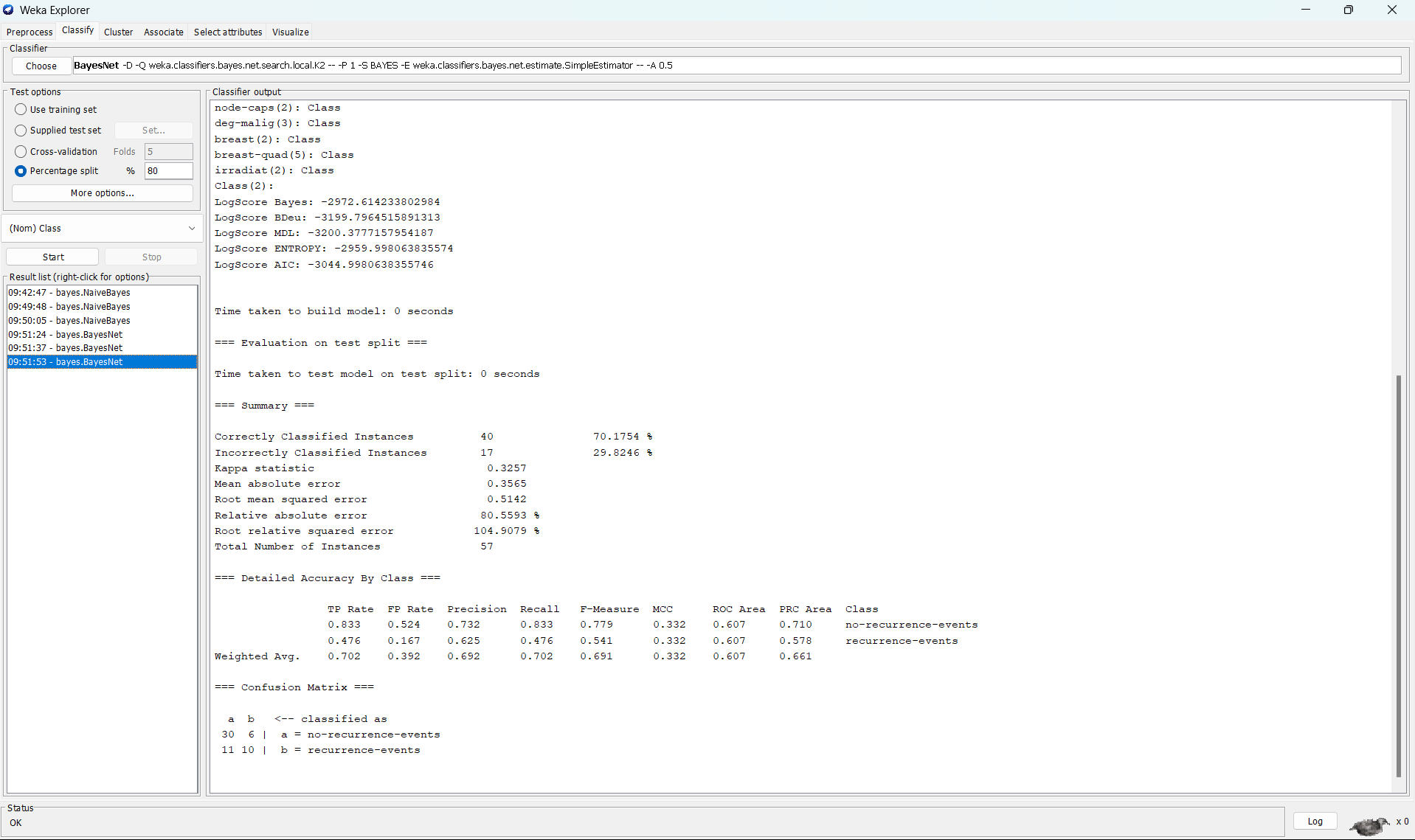
Folds(10)



Folds(5)

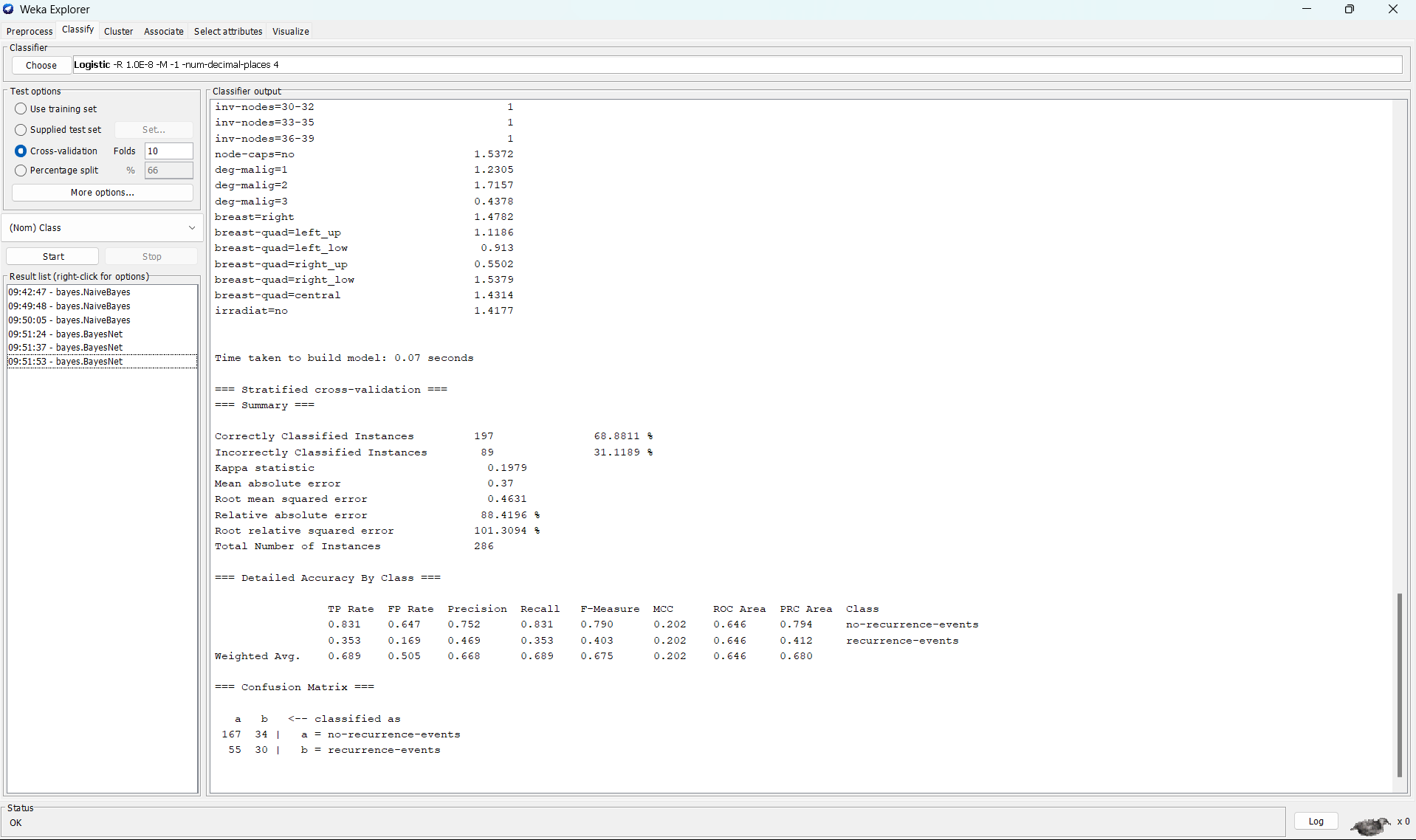


split(80%)

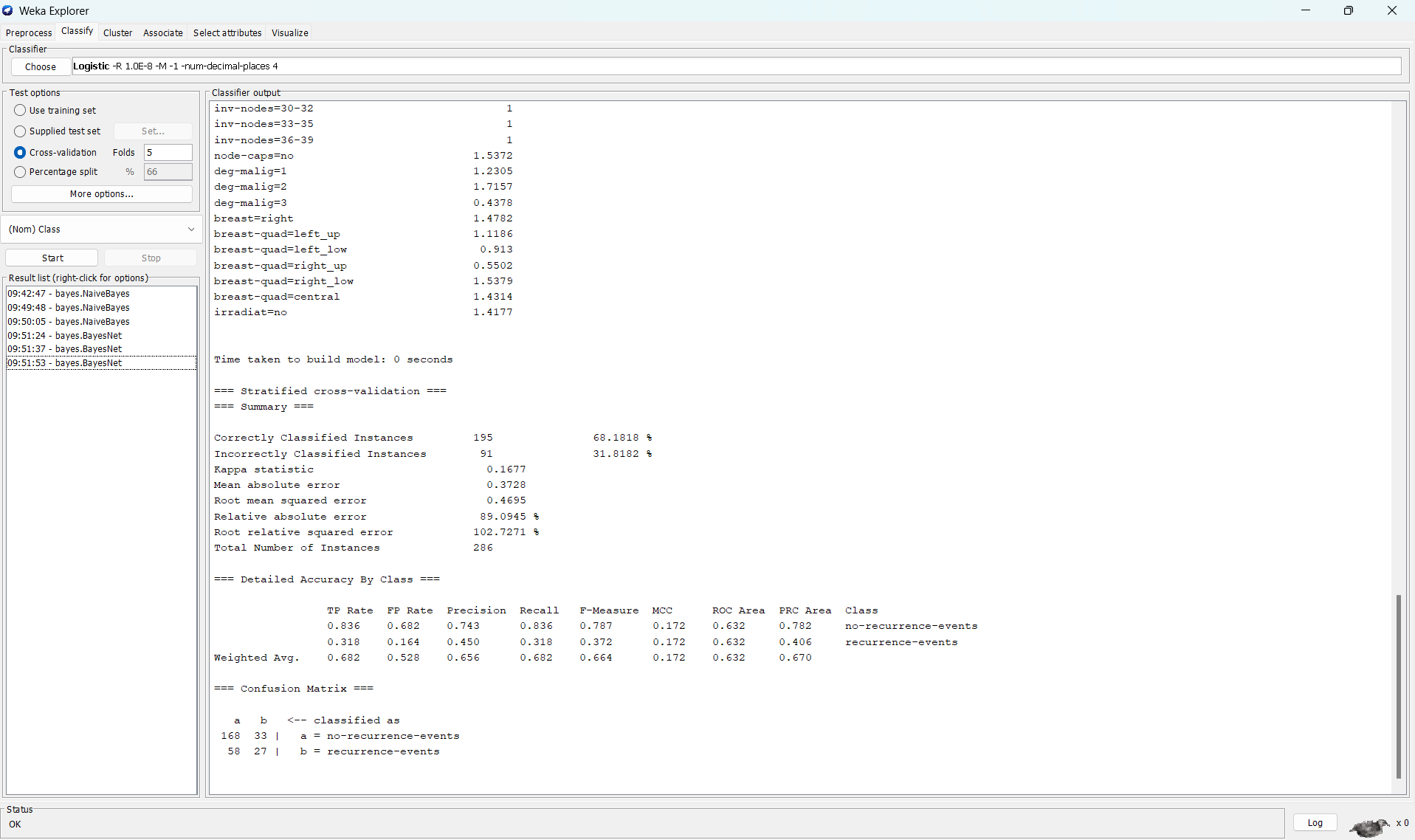


1. Logistic

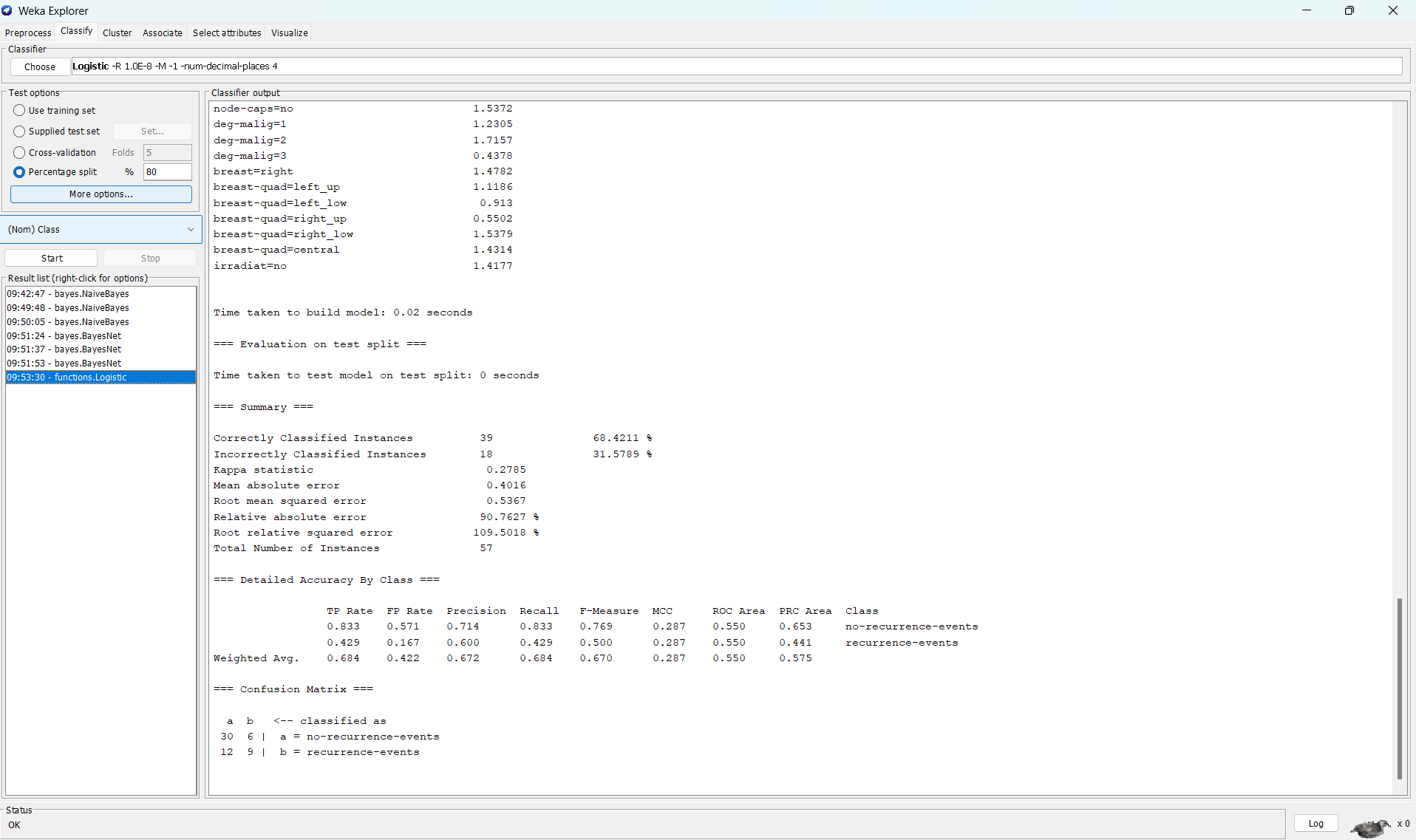
Folds(10)



Folds(5)

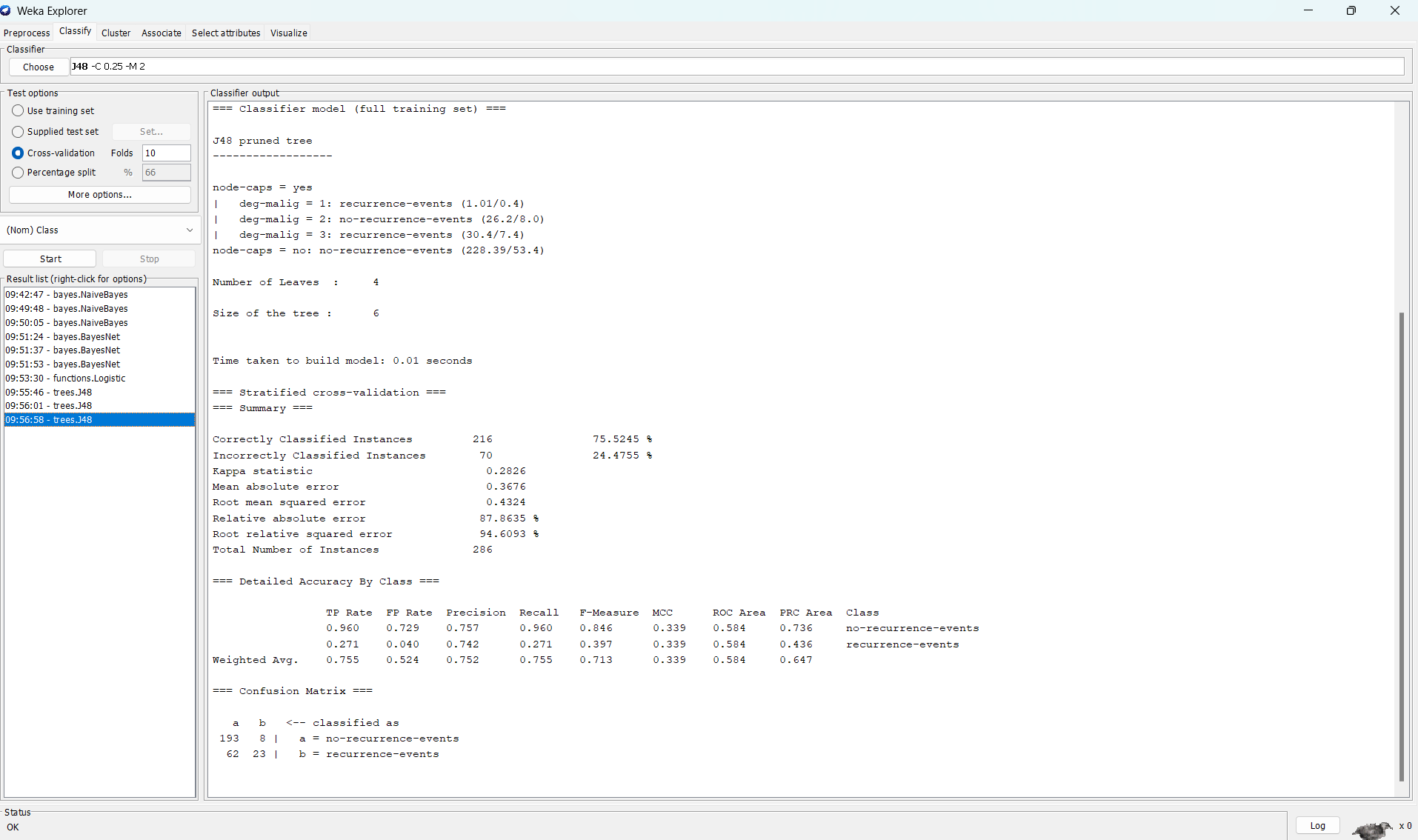


split(80%)

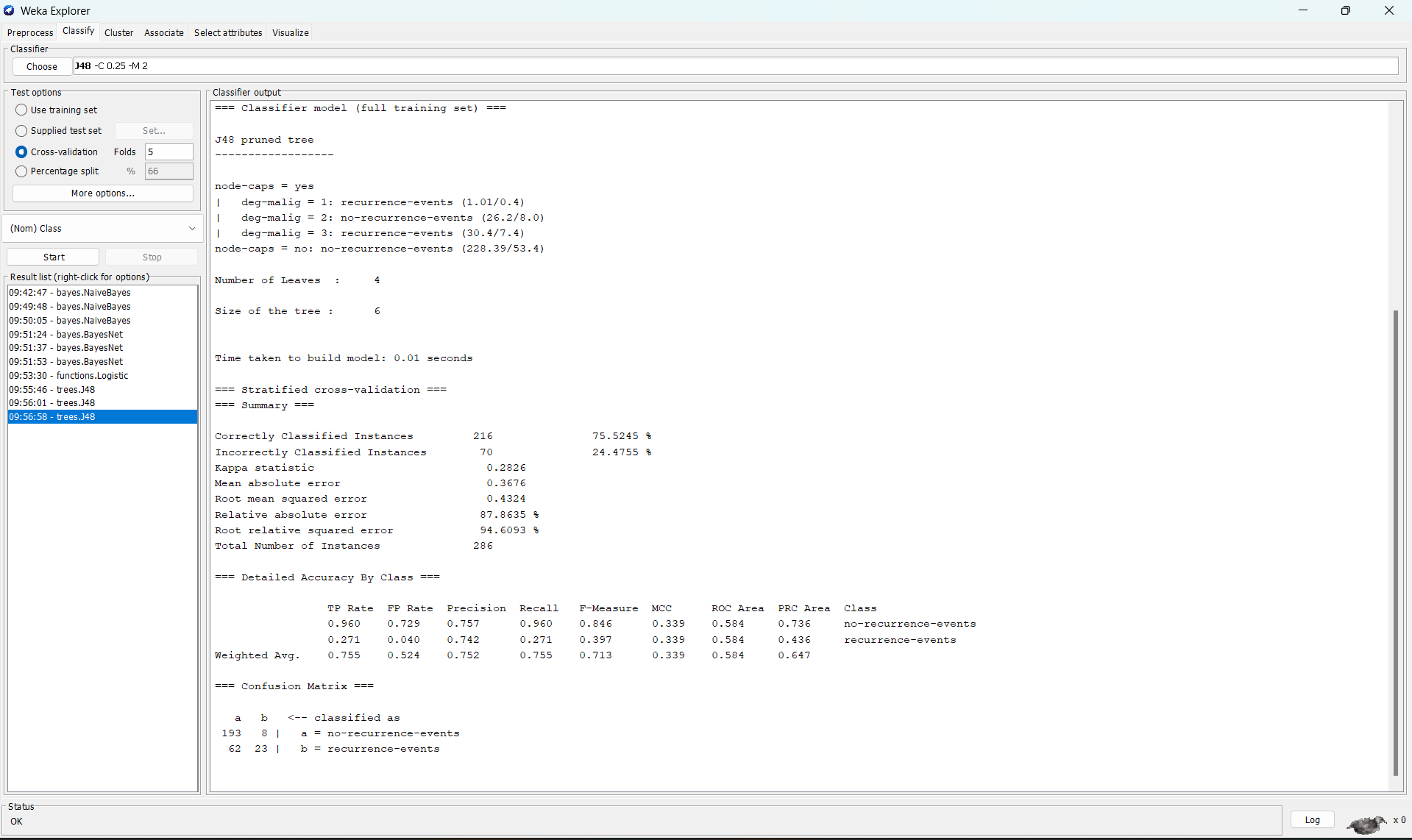


1. J48

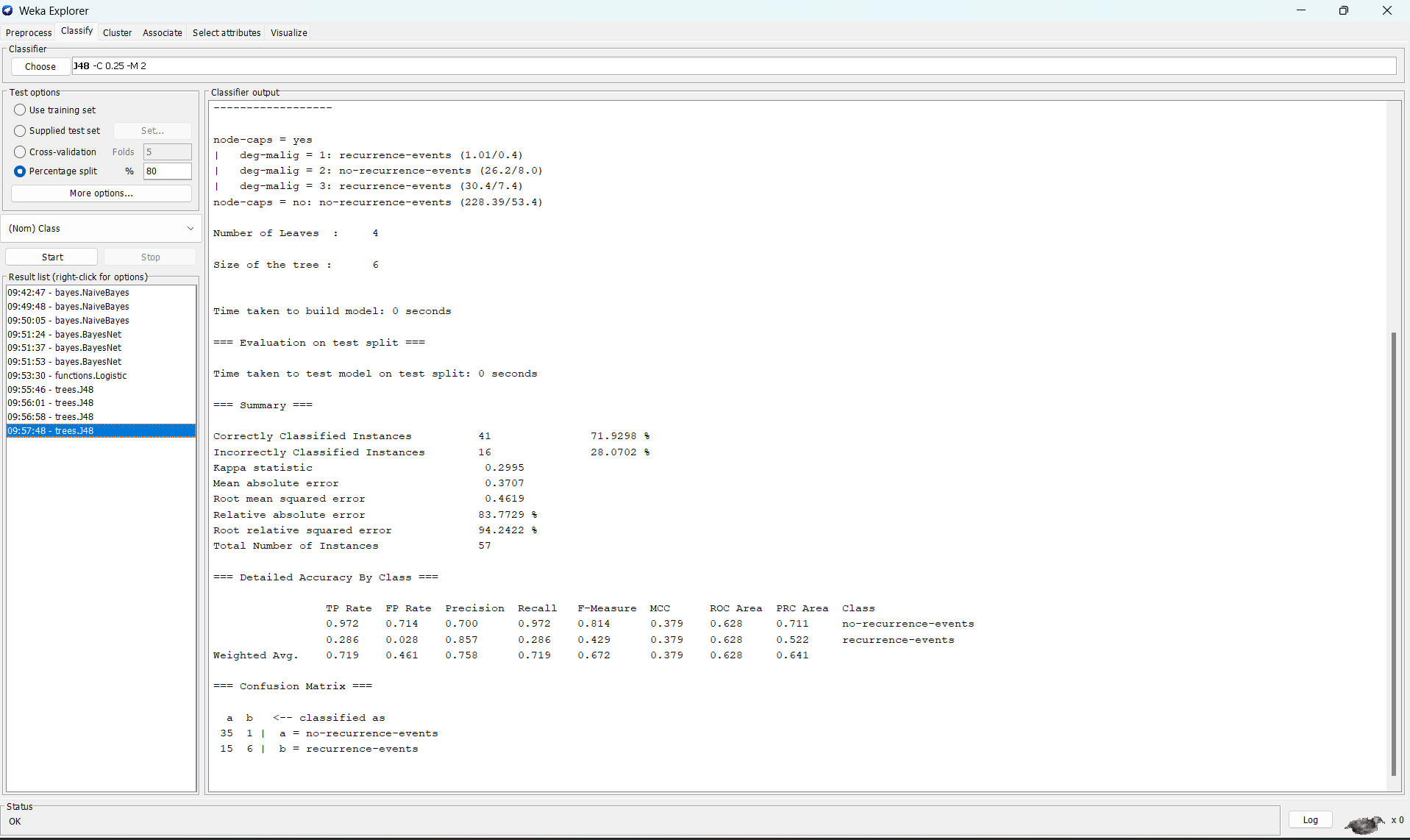
Folds(10)



Folds(5)

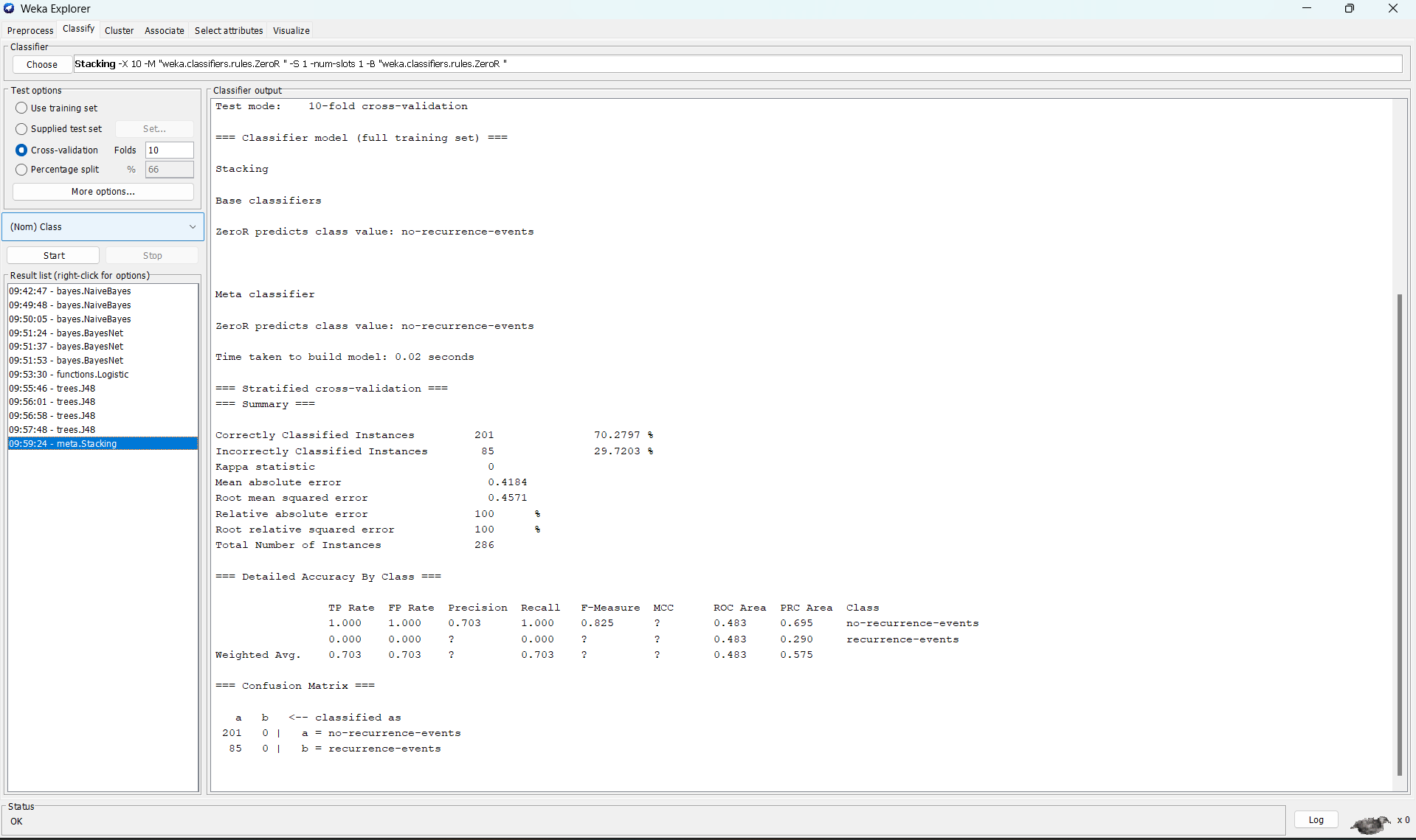


split(80%)

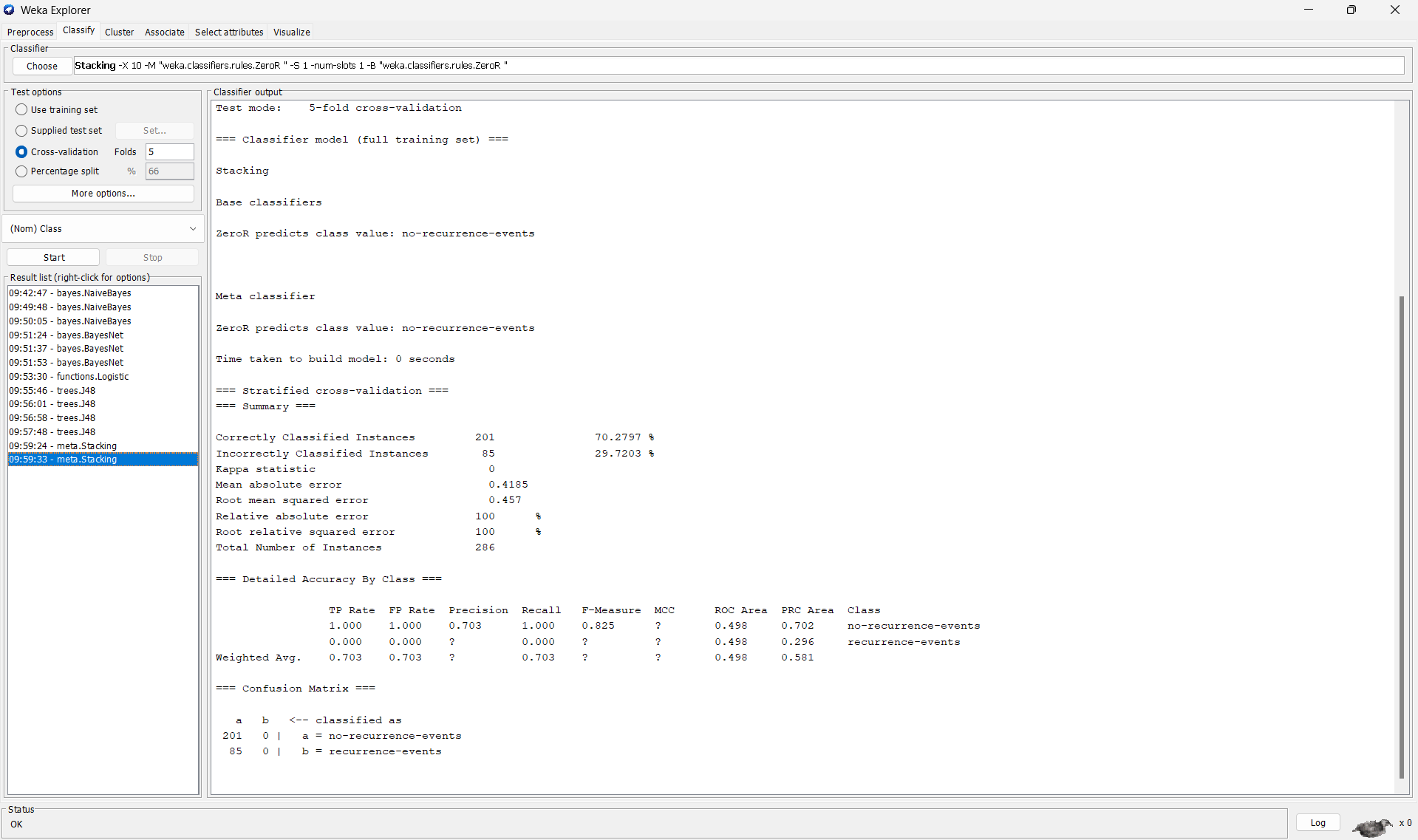


1. Stacking

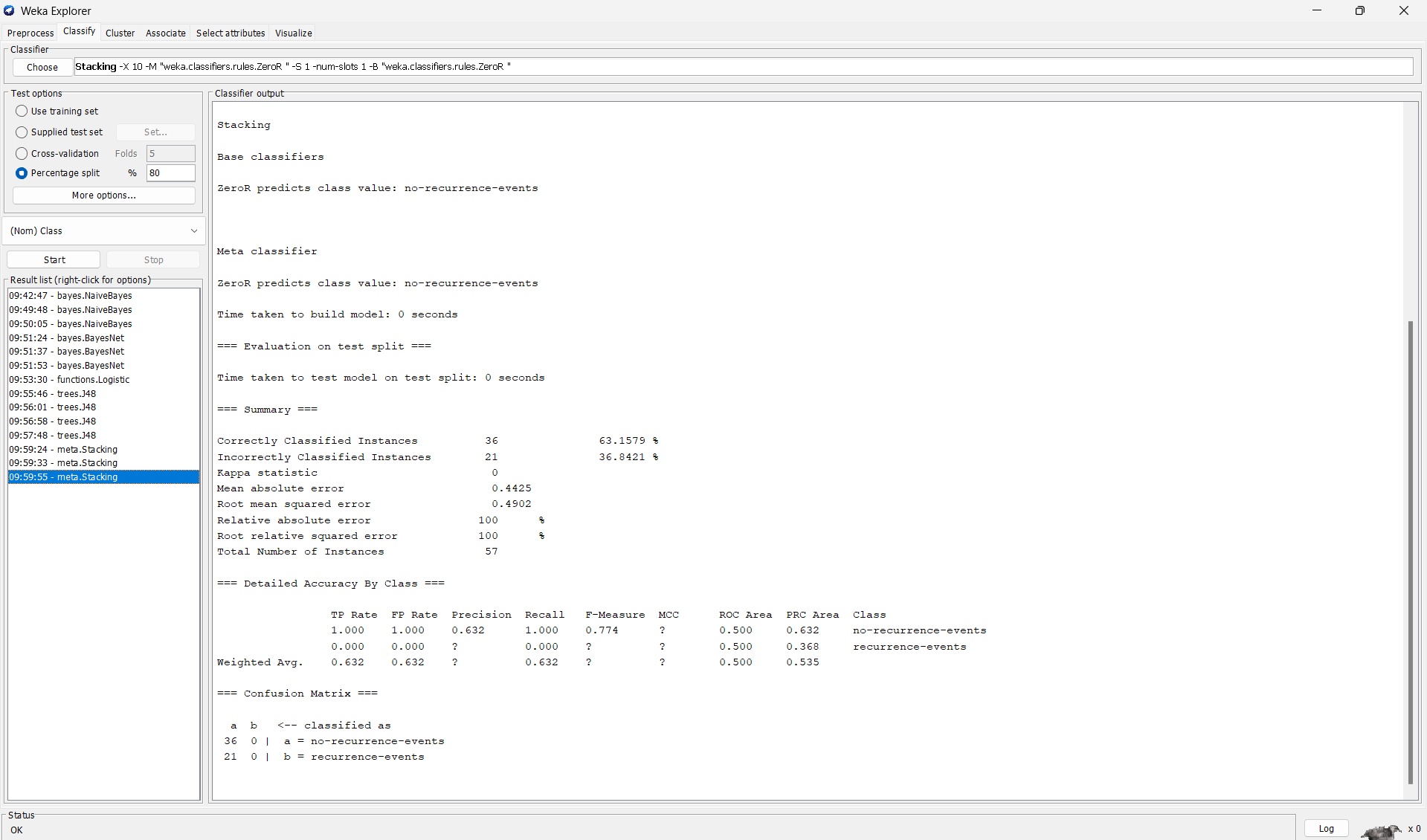
Folds(10)



Folds(5)

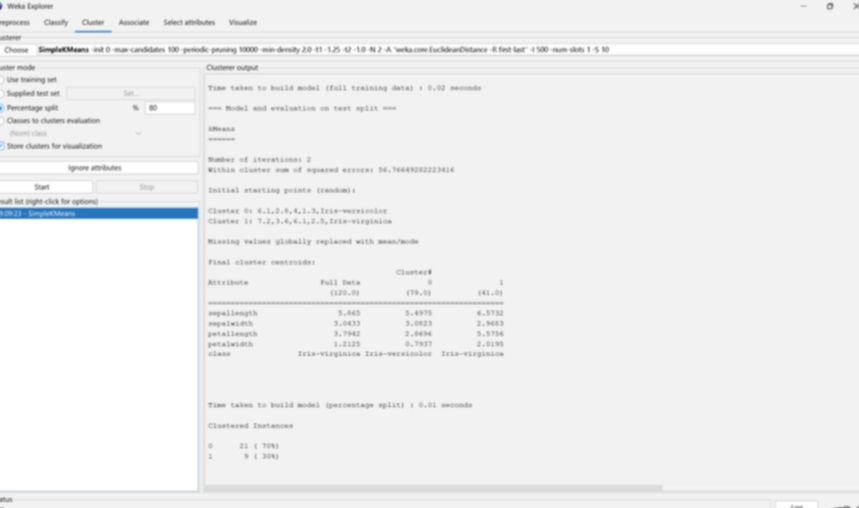


split(80%)



**Exp 3**

**Clustering using Simple K Means with Iris dataset**



• Clustering: Clustering in machine learning is an unsupervised technique that groups similar data points together based on certain traits or attributes without requiring predefined labels. It helps in identifying structures or patterns within unlabelled datasets by categorizing data into clusters of similar elements.

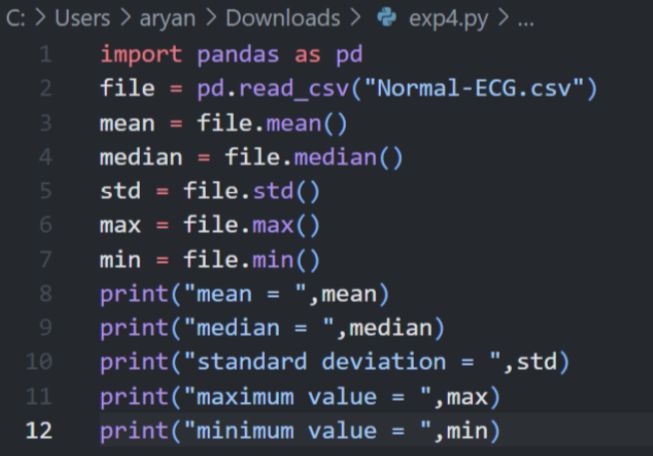
• Simple K Means: Simple K-Means is a straightforward unsupervised machine learning algorithm used for clustering. It partitions a dataset into a predefined number of clusters (denoted as K).

• Dataset Used: The Iris dataset is a classic dataset used in machine learning, particularly for classification tasks. It consists of 150 samples of iris flowers, each with four attributes: sepal length, sepal width, petal length, and petal width (all measured in centimetres). These attributes are used to classify the flowers into three species: Iris setosa, Iris versicolor, and Iris virginica, with 50 samples for each species.

**Exp 4**

**Factors to evaluate mean, median, standard deviation of data maximum, minimum**

Normal-ECG.csv



Dia-ECG.csv

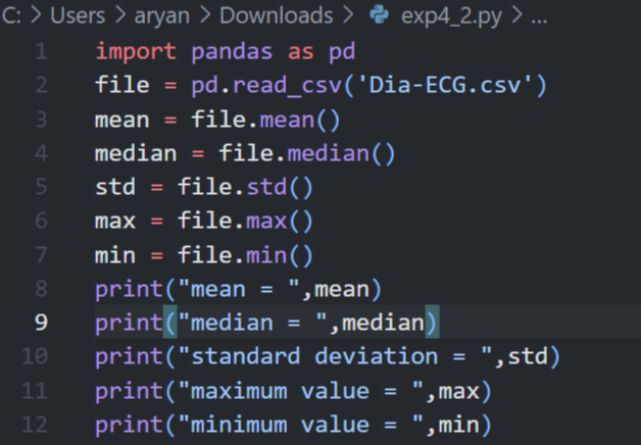


TABLE FOR COMPARISION BETWEEN NORMAL AND DIABETIC DATA SETS

