**CS7005 - BIG DATA ANALYTICS**

**Data Analysis and Visualization Using Weka Tool**

* **2017103043- Sujithra.R**

**2017103052- Vishnooprya.M**

**Dietary Patterns:**

A dietary pattern is defined as the quantity, variety, or combination of different foods and beverage in a diet and the frequency with which they are habitually consumed.

The basic dietary patterns are:

* **Vegetarians:** Those who eat vegetables and dairy products only.
* **Lacto-ovo** **vegetarian:** Some may also consider eggs to be a vegetarian diet.
* **Non-vegetarians:** Those who have vegetables, meat, eggs and dairy products as a part of their diet.
* **Vegans:** Those who have only vegetables and plant-derived foods for their meals.

**Analysing Dietary Patterns:**

The dietary patterns of people around us are obtained and are analysed and classified based on various factors like age, gender, BMI, exercise patterns, vitamin needs, habit of eating junk foods, usage of other animal products (like leather), price of the foods, health consciousness, moral values and culture and the reasons that make them follow a particular diet pattern.

**Data Collection:**

The data is collected through a survey conducted using google forms among college students, most of them between the age of 18-21. This obtained data can be used for ideas to introduce new food in the college mess or canteen.

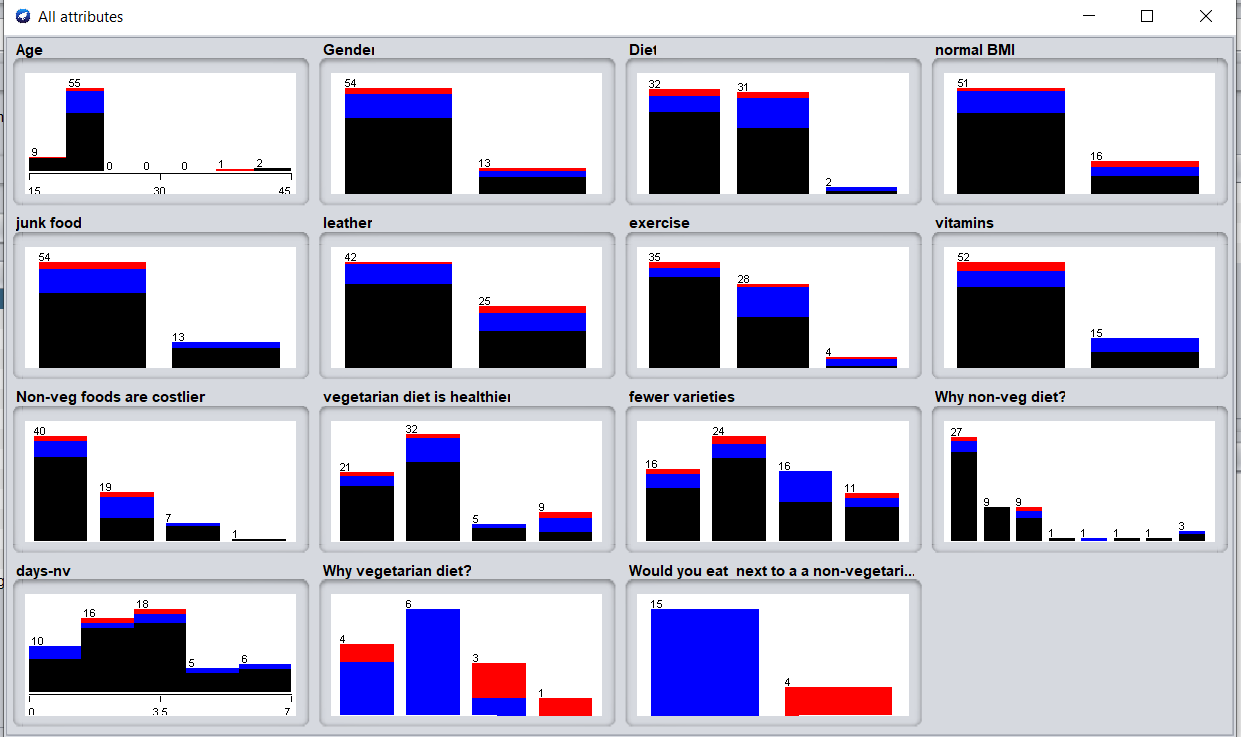
**Participants**:

67 participants were included for model evaluation.

**Classification:**

The data obtained is analysed using Weka – a data mining tool which helps us predict the patterns in the dietary habits of people and the obtained analysis can be used in various ways for formulating consumer targeting strategies.

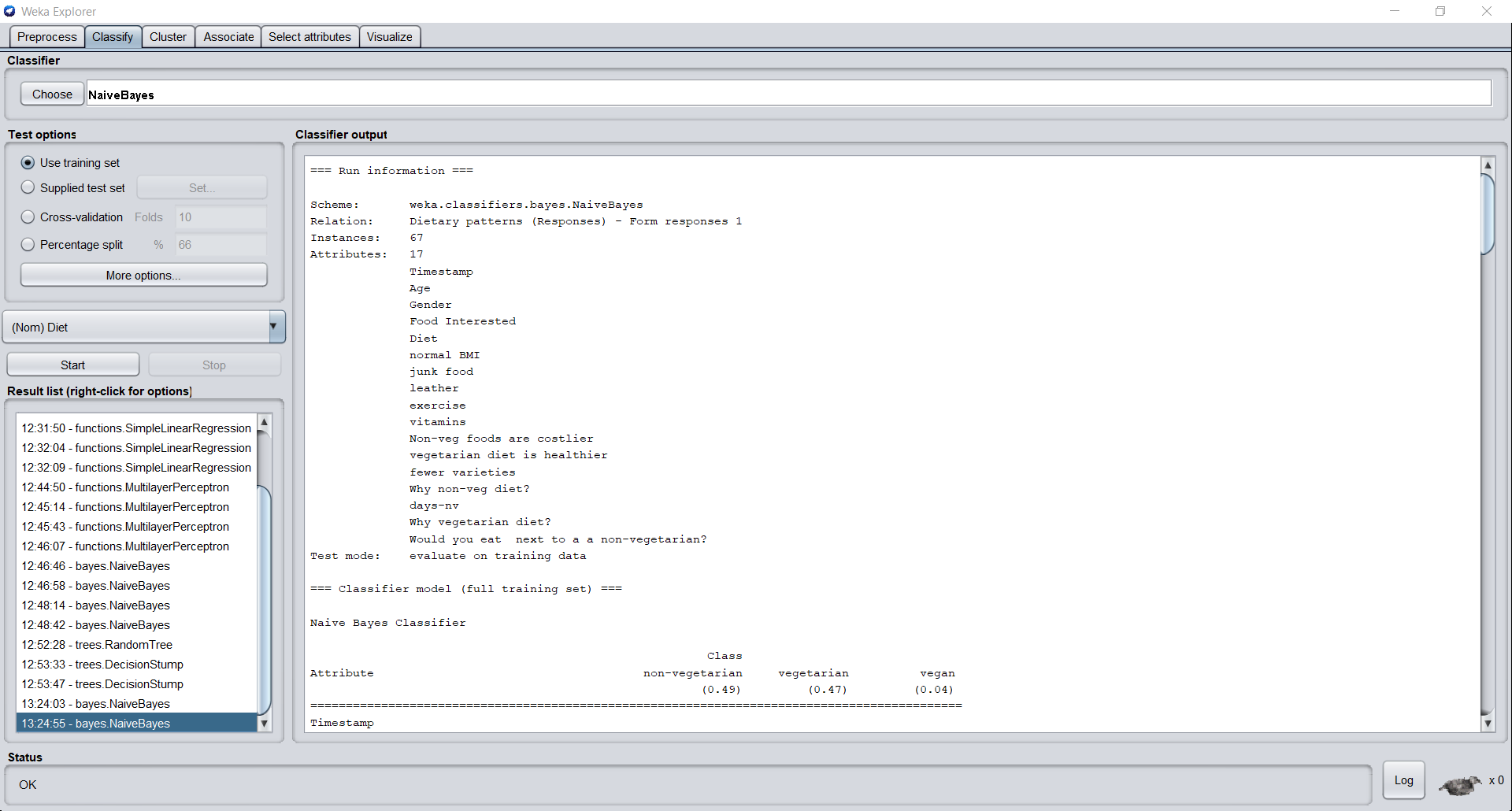
Screenshot of visualization of all attributes:

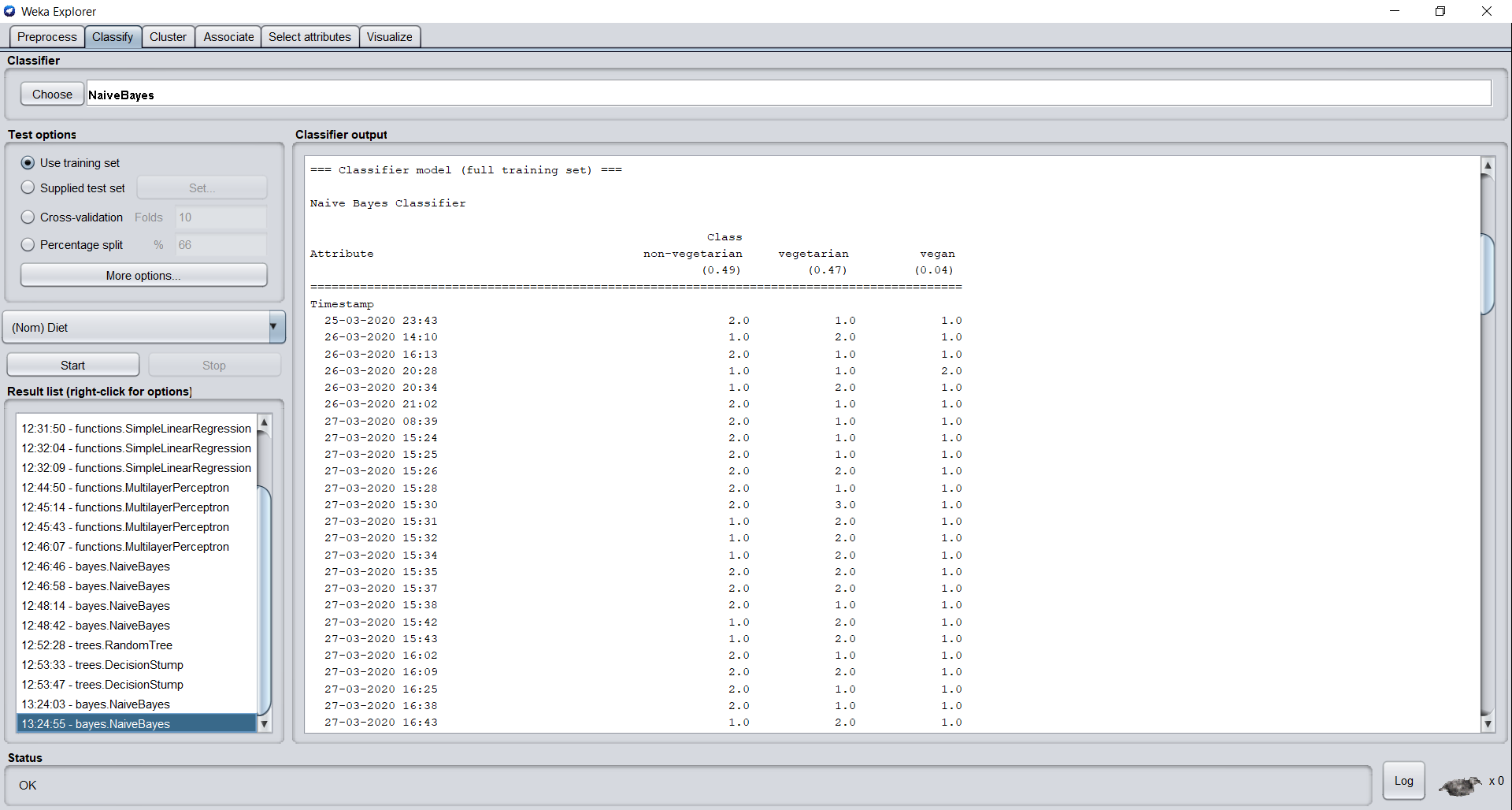


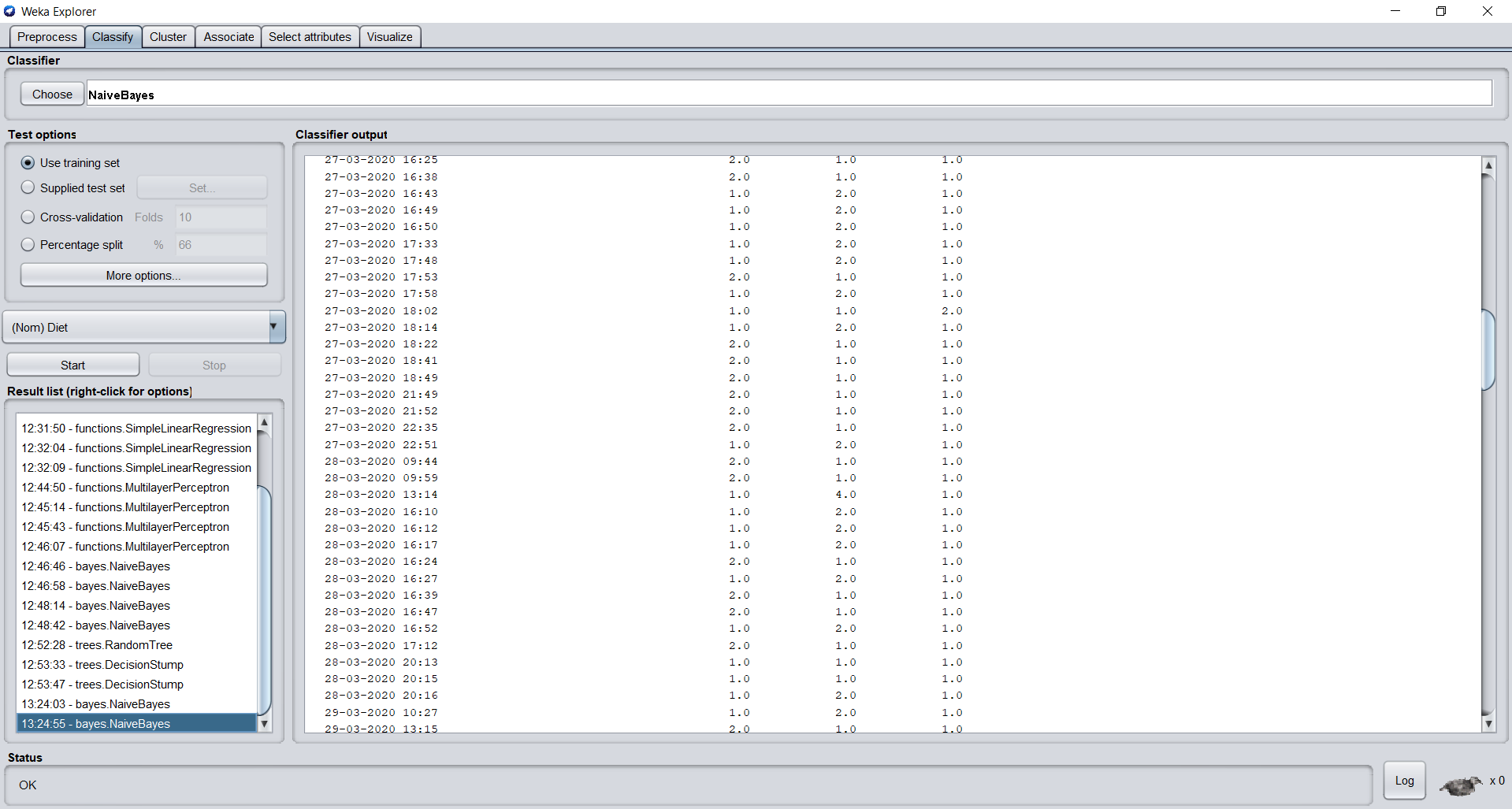
**Naive Bayes Classifier:**

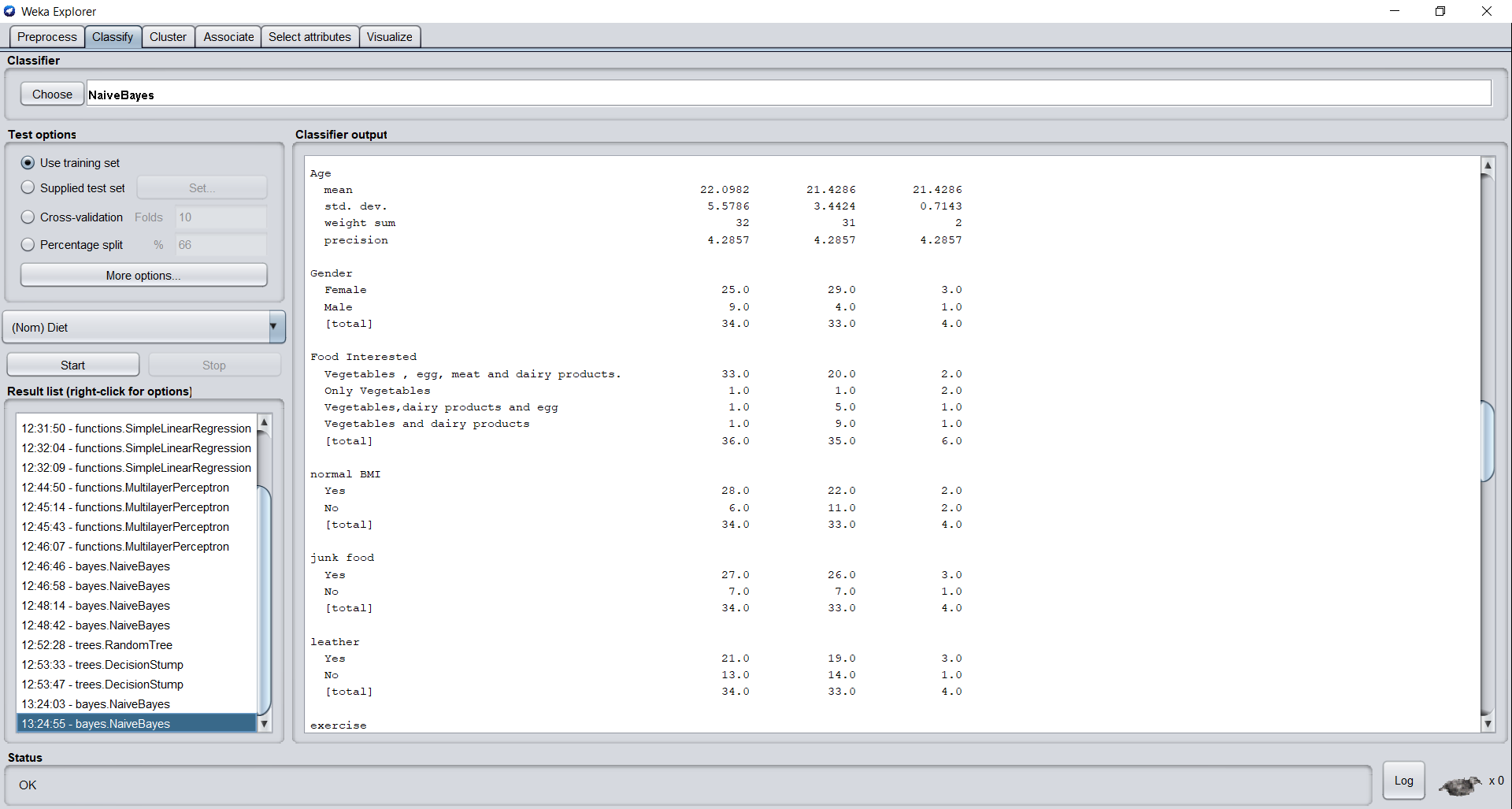
Naive Bayes classifiers are a collection of classification algorithms based on Bayes’ Theorem. It is a family of algorithms which assumes that every pair of features being classified is independent of each other.

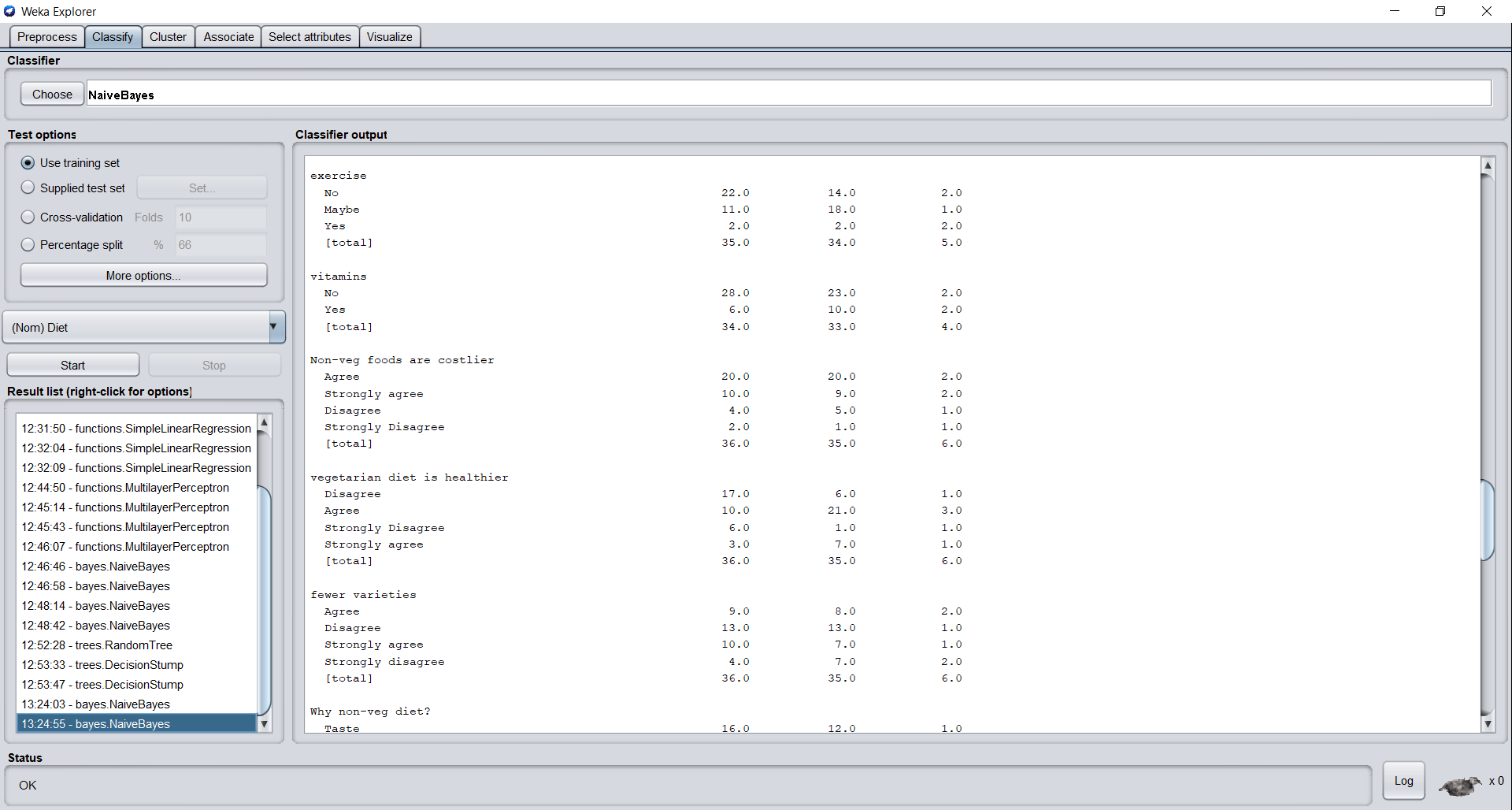
The screenshots here show the RESULT obtained:

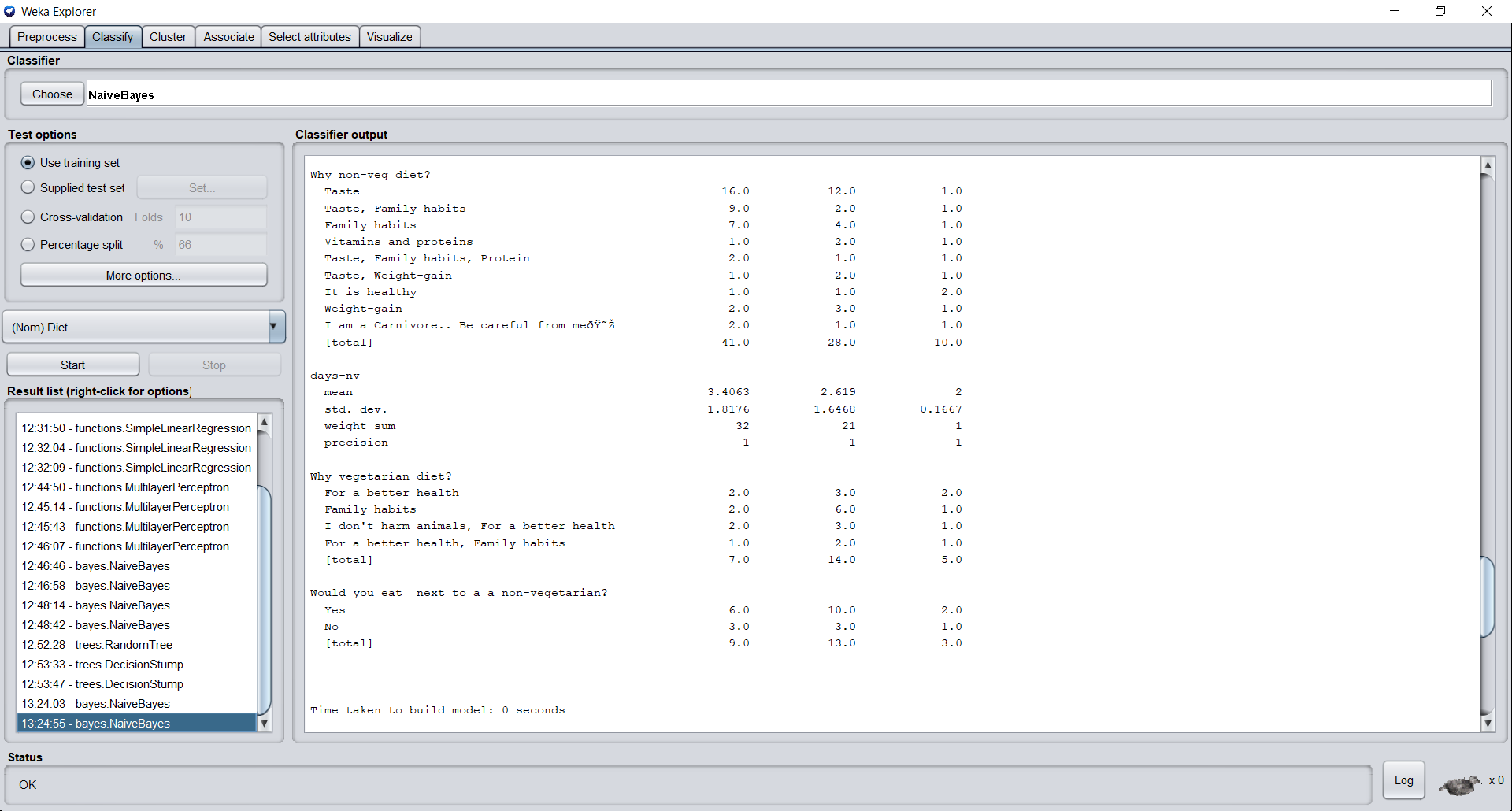


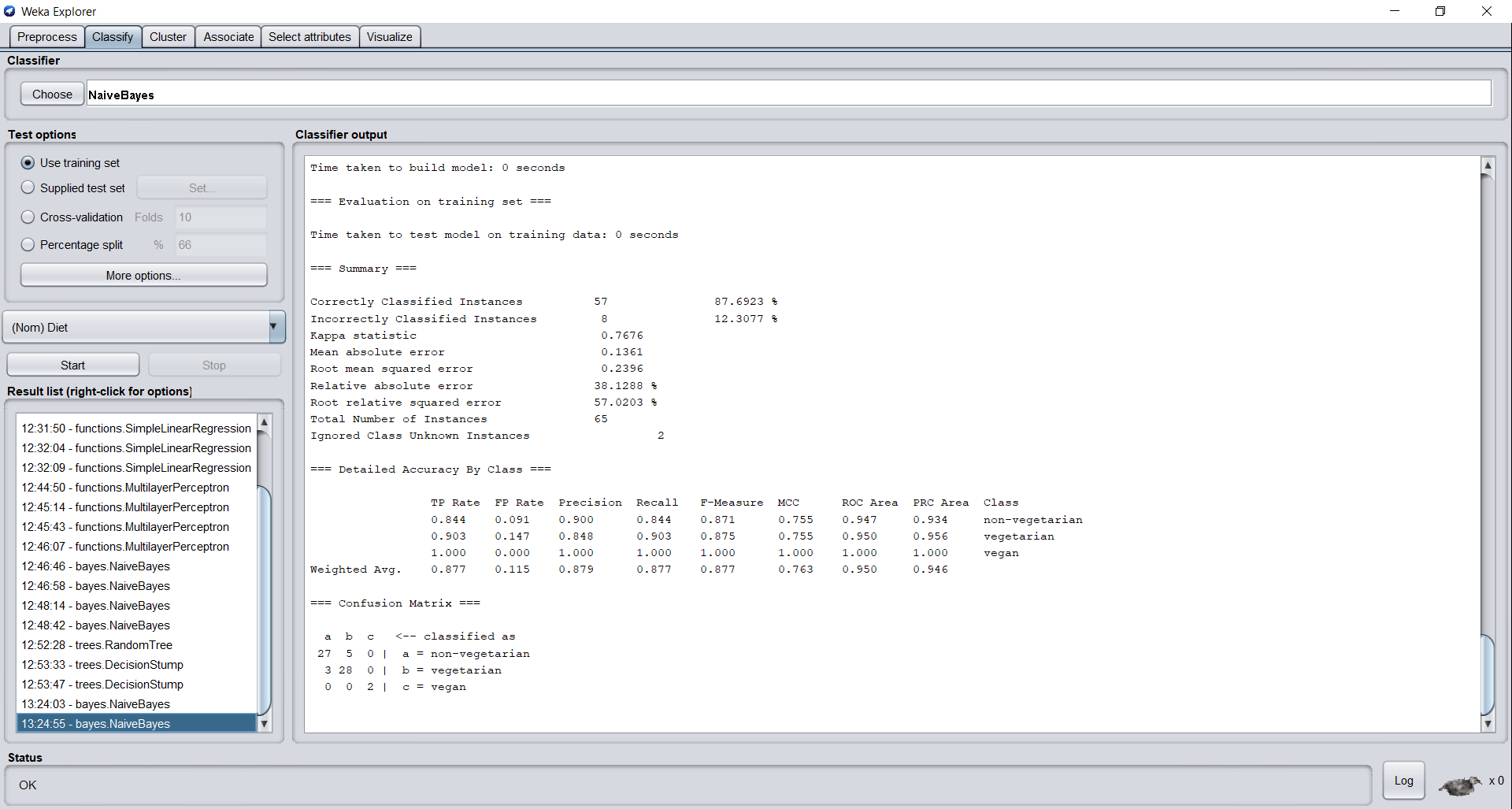




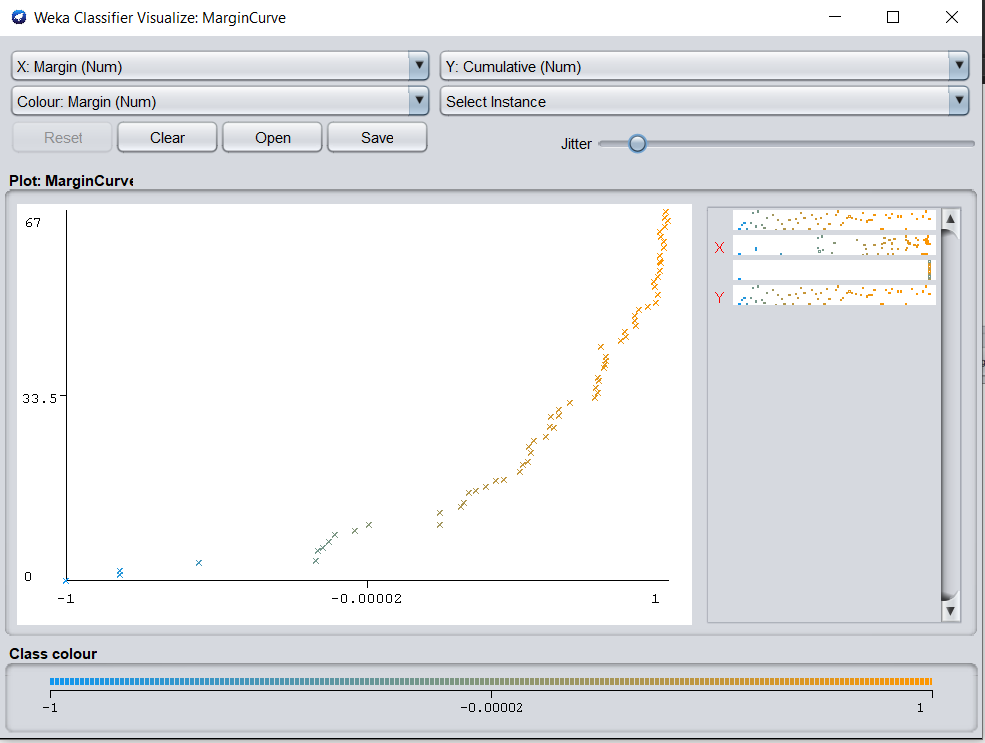
The analysis for each attribute:





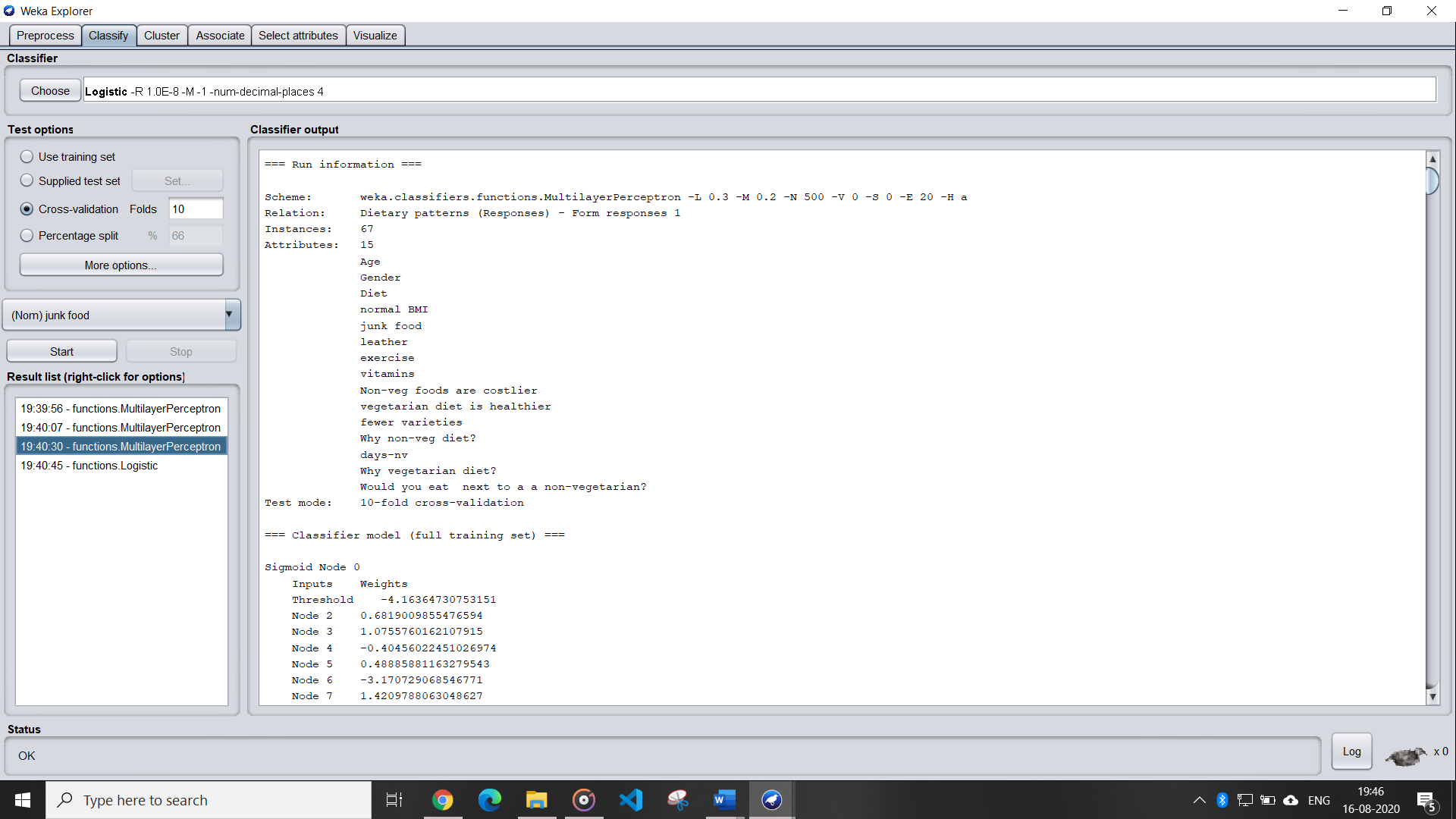
The final results are obtained as:

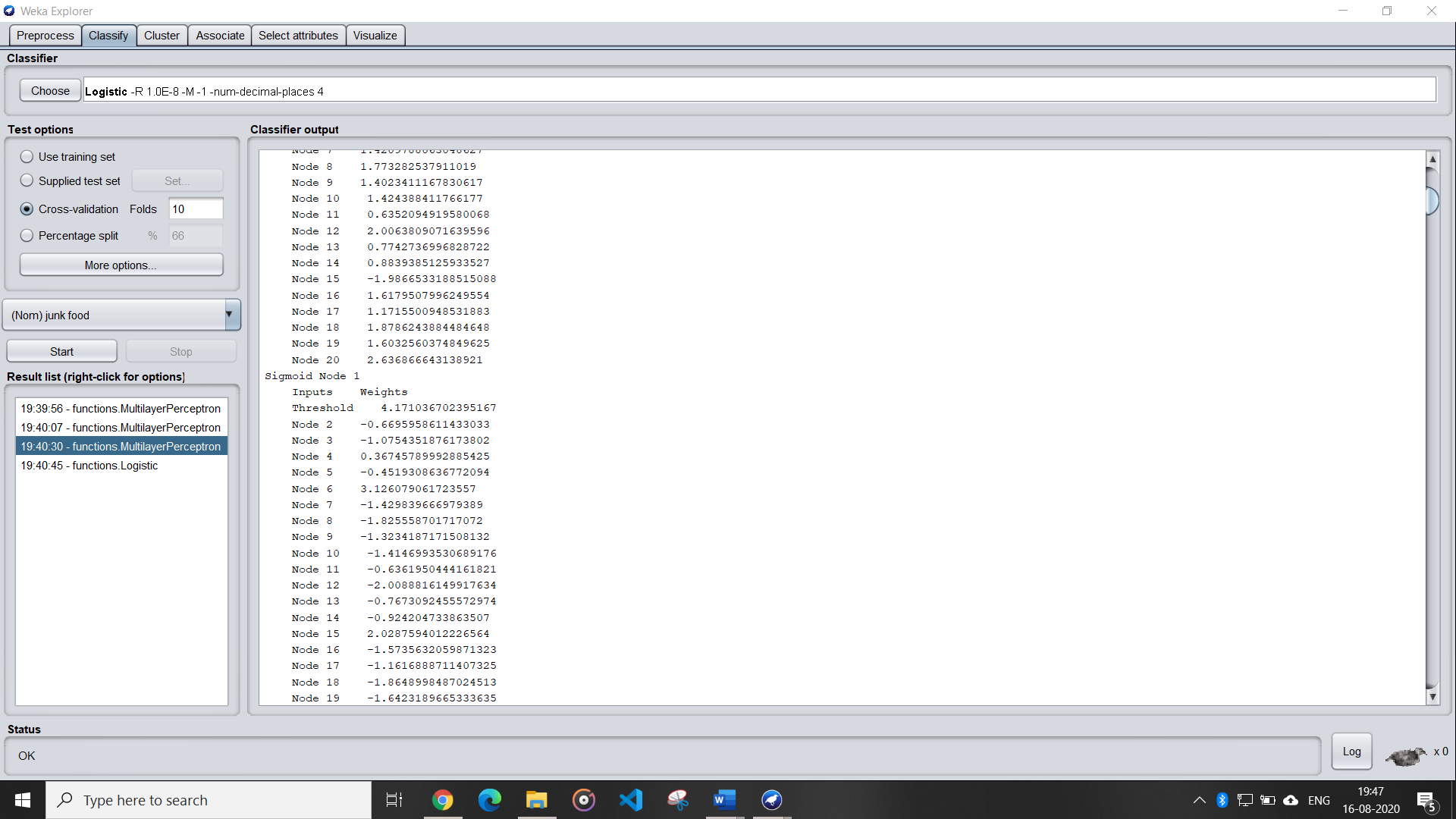
Visualization of the classification:

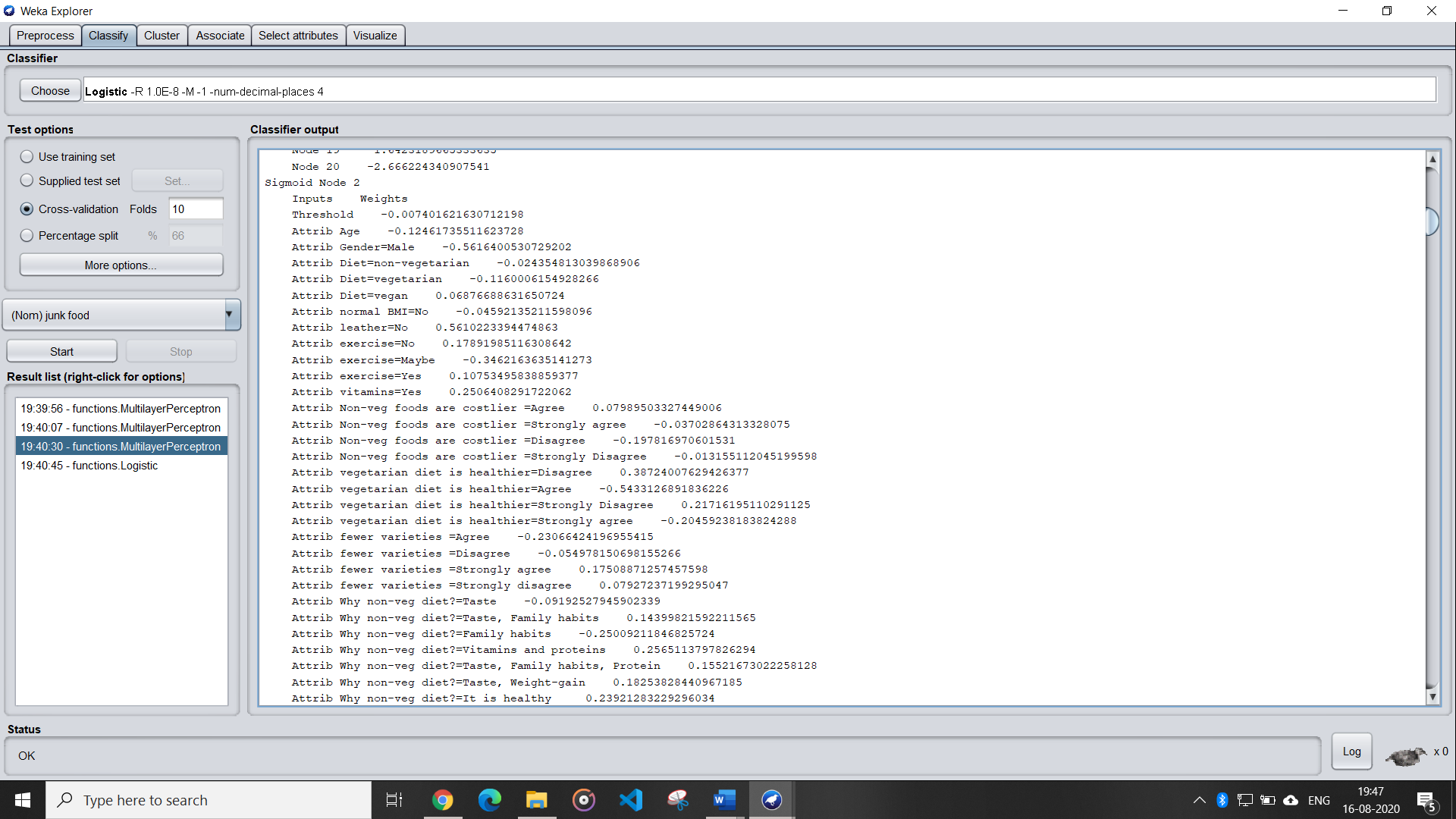


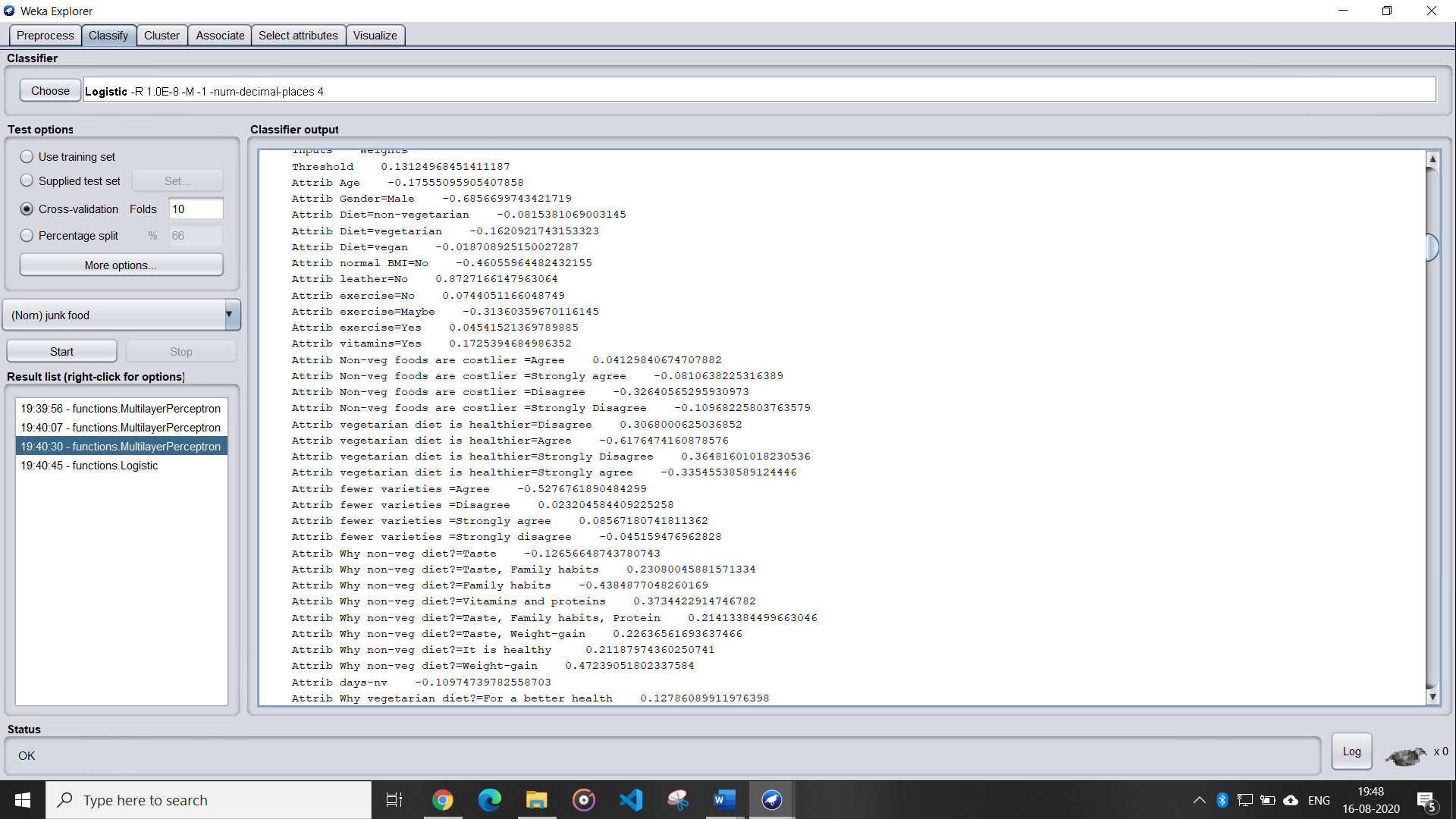
**Multilayer Perceptron:**

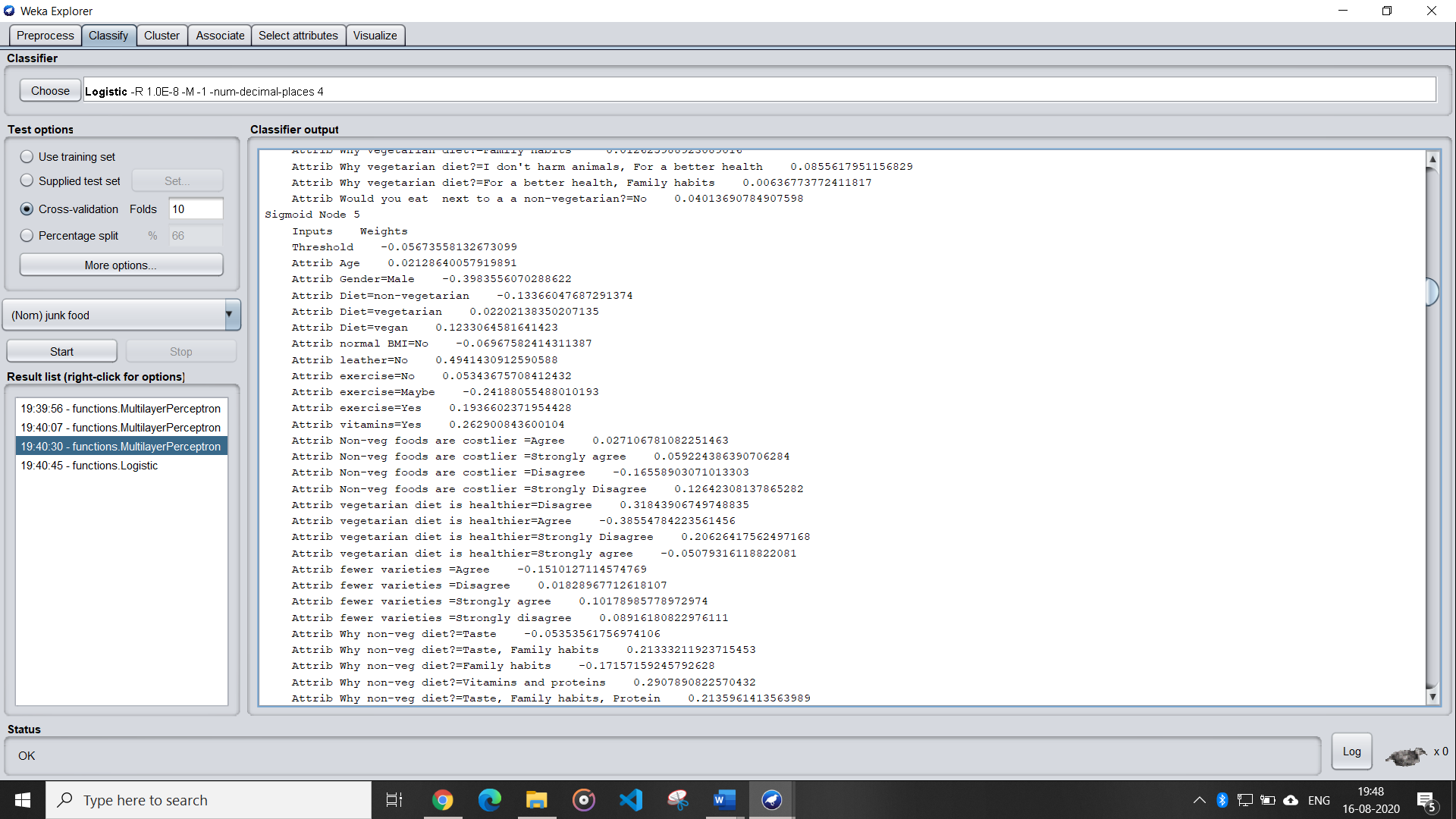
A **multilayer perceptron** (MLP) is a class of [feedforward](https://en.wikipedia.org/wiki/Feedforward_neural_network) [artificial neural network](https://en.wikipedia.org/wiki/Artificial_neural_network) (ANN). The term MLP is used ambiguously, sometimes loosely to any feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of [perceptron](https://en.wikipedia.org/wiki/Perceptron).



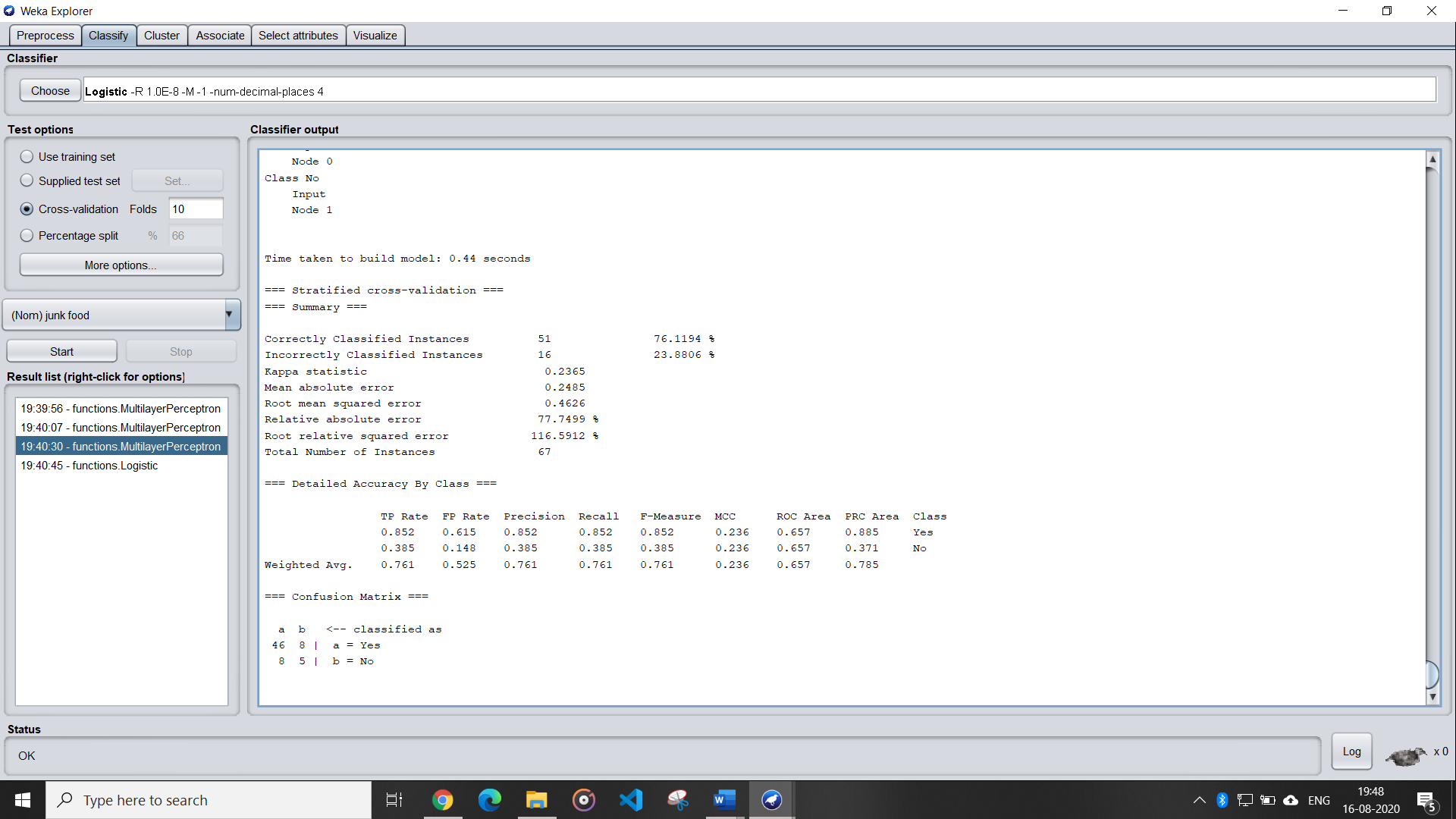






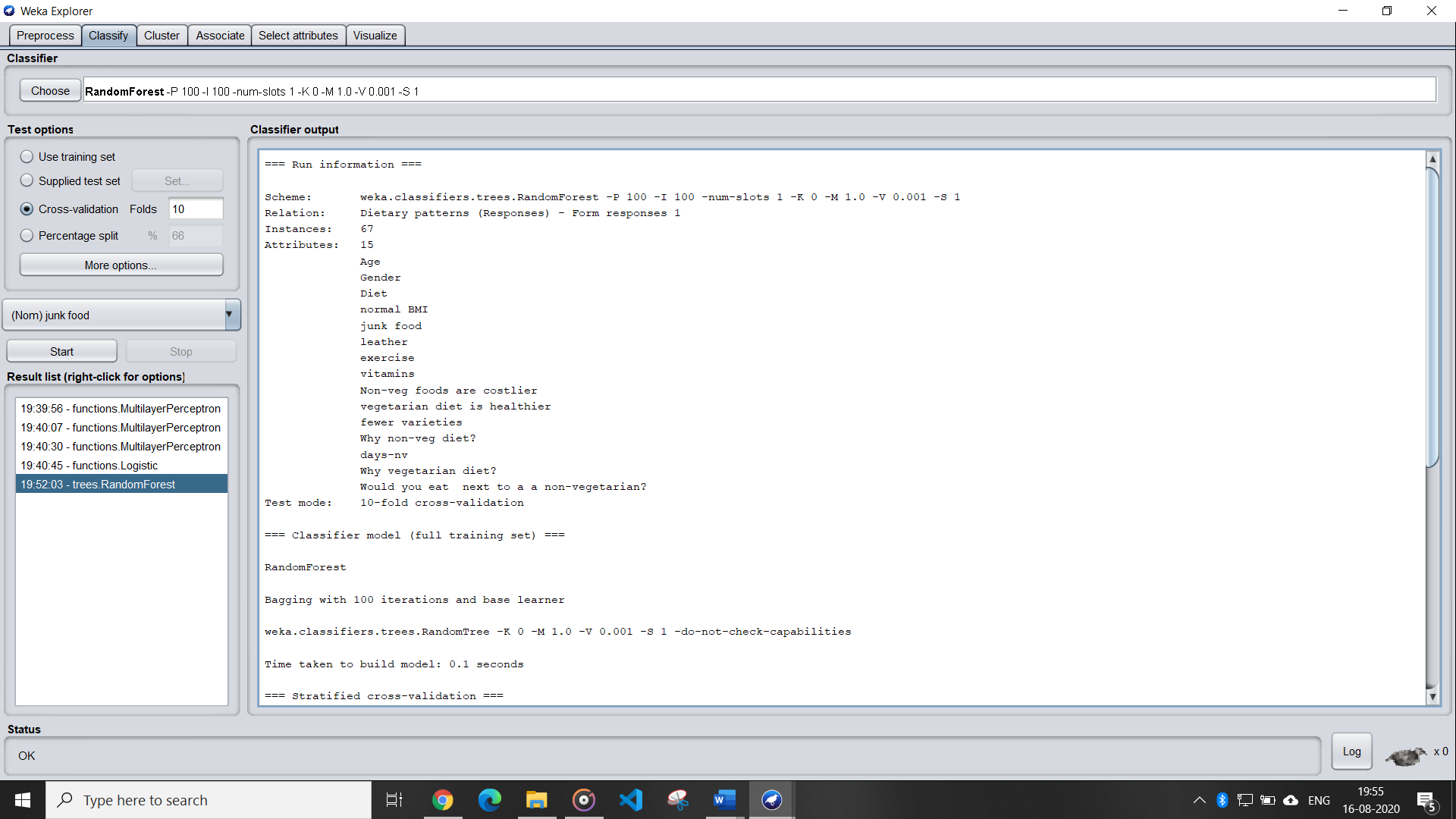


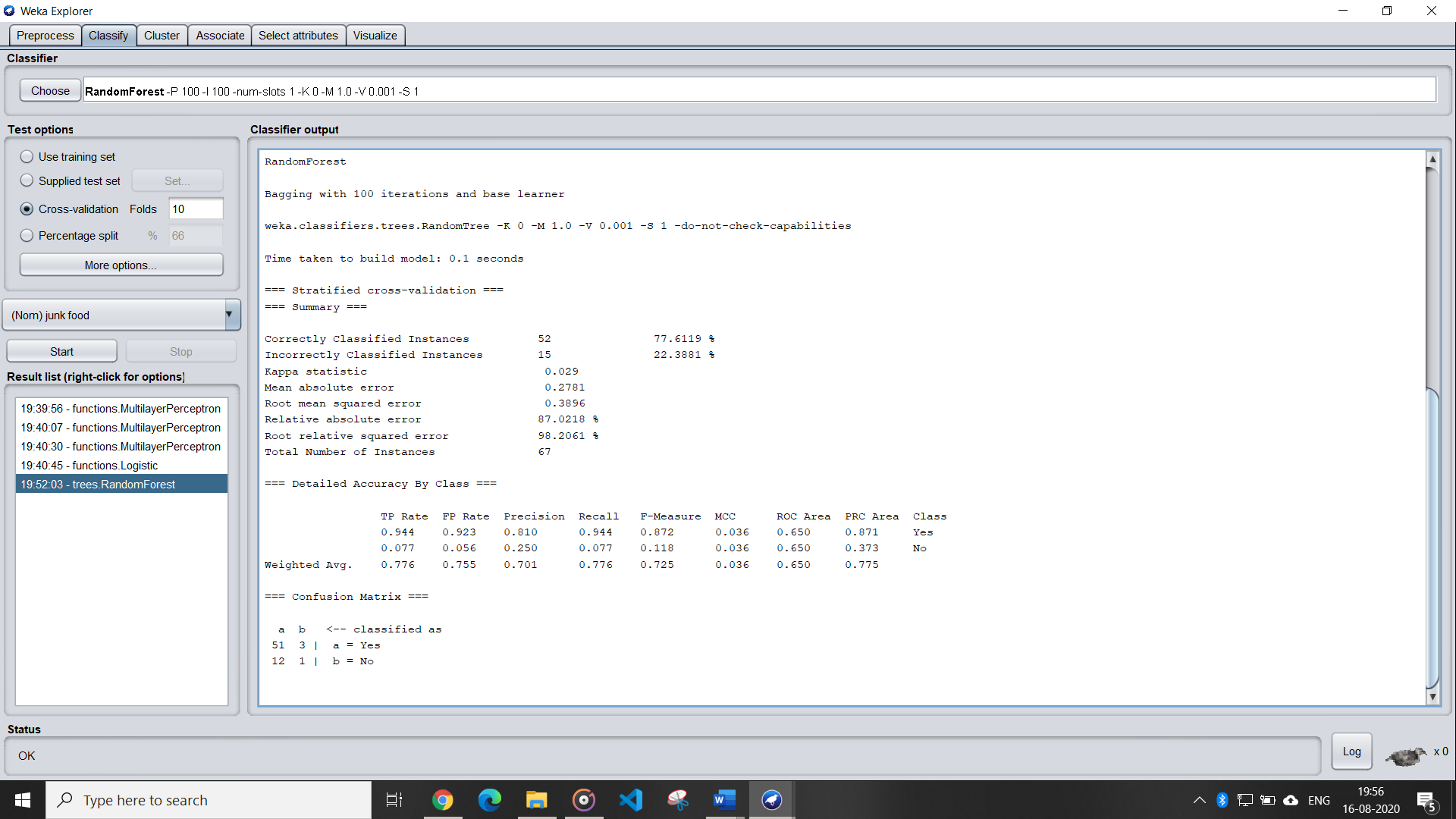
The final results are:



**Random Forest:**

Random forest is a supervised learning algorithm which is used for both classification as well as regression. Random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution.





**Simple K-means clustering:**

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

