**CST8390 - Lab 4**

**Classification by Decision Trees**

**Student Name: Onur Onel**

**Student Number: 041074824**

a. Number of instances: **768**

b. Number of attributes: **8 plus the class attribute, totaling 9 attributes.**

c. List of attributes:

* **Number of times pregnant**
* **Plasma glucose concentration a 2 hours in an oral glucose tolerance test**
* **Diastolic blood pressure (mm Hg)**
* **Triceps skin fold thickness (mm)**
* **2-Hour serum insulin (mu U/ml)**
* **Body mass index (weight in kg/(height in m)^2)**
* **Diabetes pedigree function**
* **Age (years)**
* **Class (tested\_negative or tested\_positive)**

d. Class labels and their relabelled values:

* **Class label 0: tested\_negative**
* **Class label 1: tested\_positive**

e. Number of instances for each class label:

* **tested\_negative: 500 instances**
* **tested\_positive: 268 instances**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Confusion Matrix** | **Number of leaves** | **Size of the tree** | **Correctly classified instances** |
| 5 | **a b <-- classified as**  **417 83 | a = tested\_negative**  **100 168 | b = tested\_positive** | **20** | **39** | **585** |
| 7 | **a b <-- classified as**  **408 92 | a = tested\_negative**  **107 161 | b = tested\_positive** | **14** | **27** | **569** |
| 9 | **a b <-- classified as**  **414 86 | a = tested\_negative**  **98 170 | b = tested\_positive** | **17** | **33** | **584** |
| 10 | **a b <-- classified as**  **418 82 | a = tested\_negative**  **112 156 | b = tested\_positive** | **8** | **15** | **574** |

12.

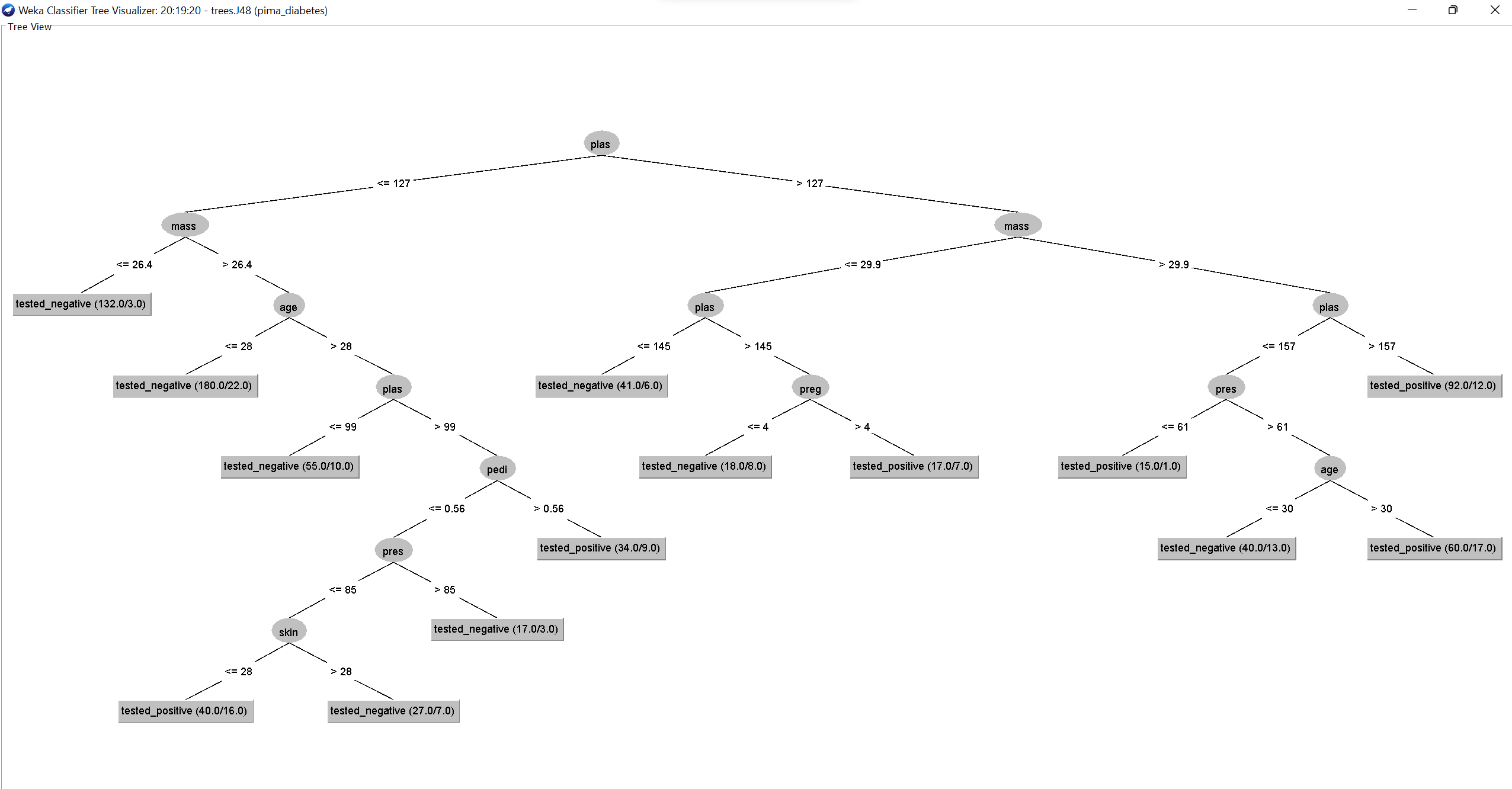
**a. Details of random forest: RandomForest with 100 iterations.**

**b. Time taken to build model: 0.18 seconds.**

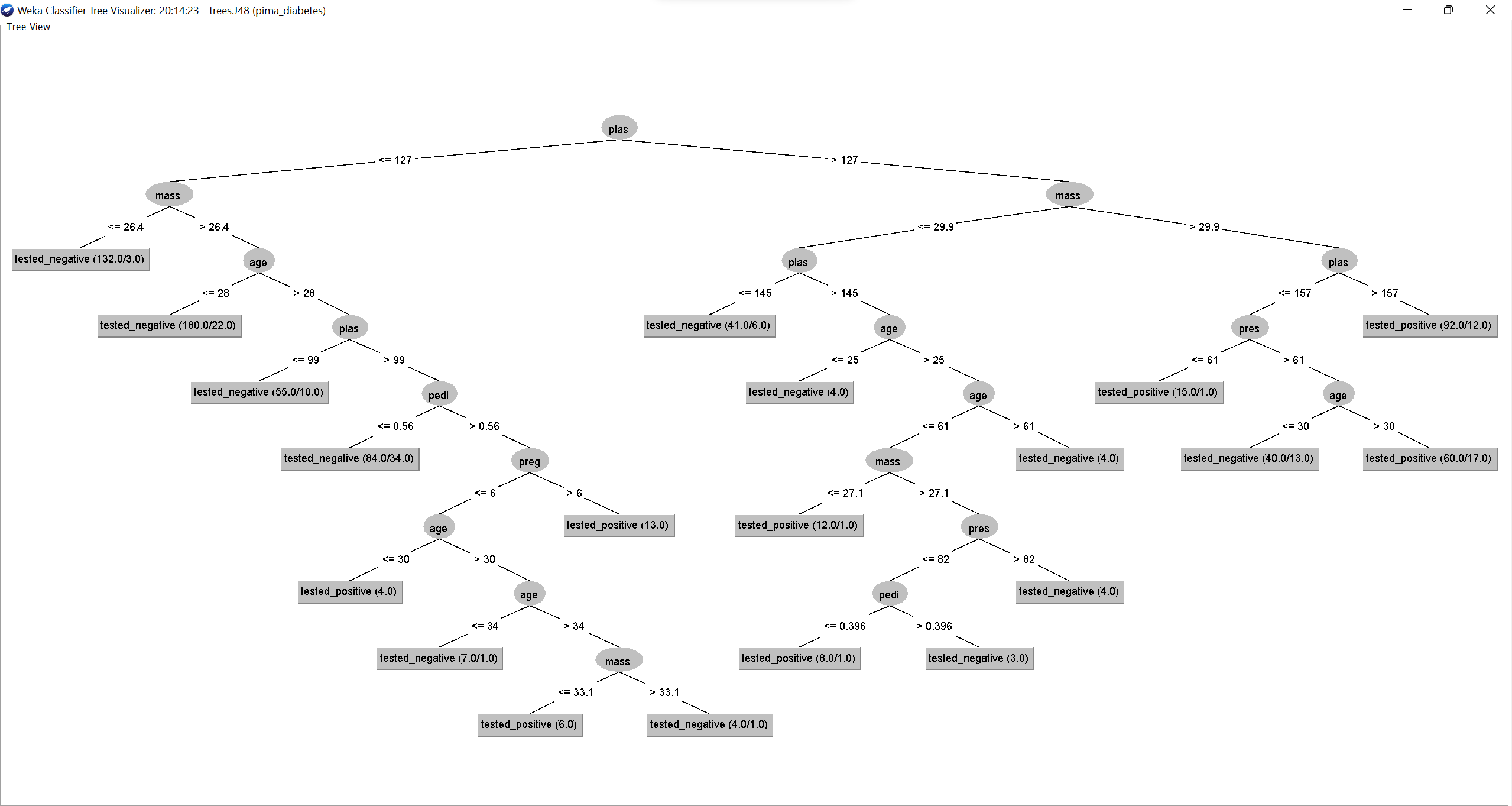
**c. Correctly classified instances: 577.**

**Screenshots with Question Number**

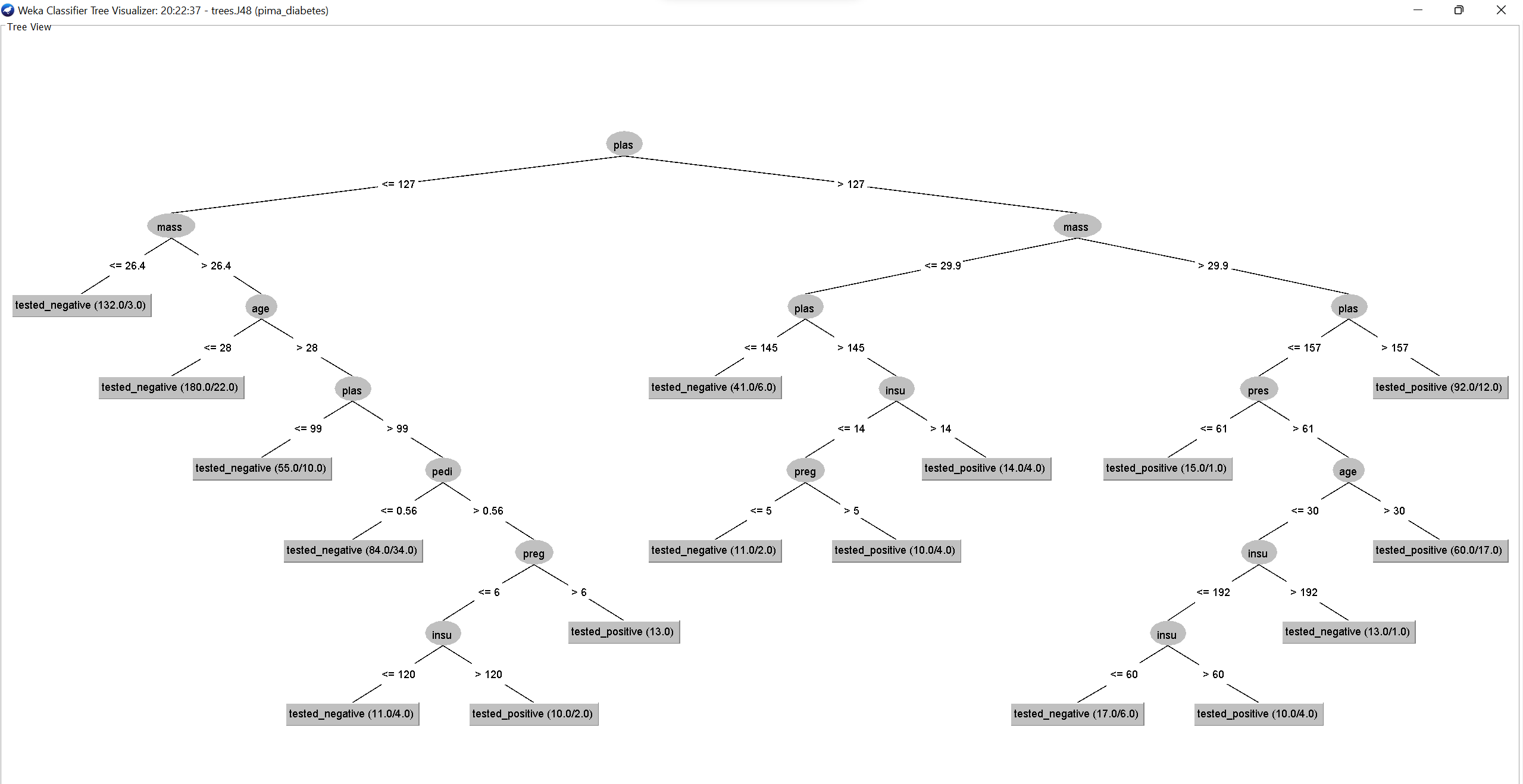
**1.**

****

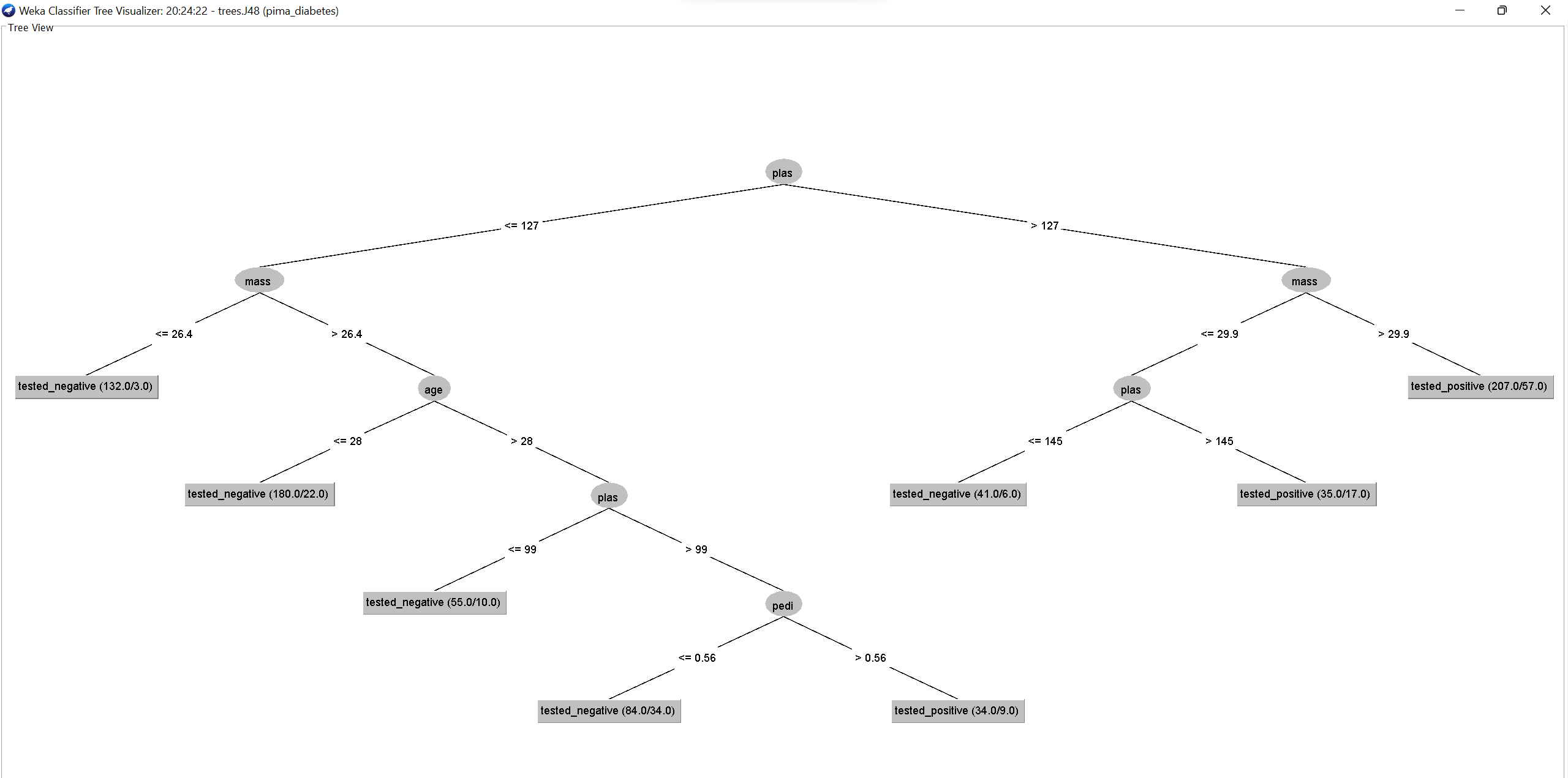
**2.**

****

**3.**

****

**4.**

****