**CST8390 - Lab 4**

**Classification by Decision Trees**

**Student Name: Neil Kingdom**

**Student Number: 040967309**

2.   
 a. Without removing duplicates, there are 768 instances

b. There are 8 attributes (plus class = 9)

c.

* 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 variable (0 or 1)

d.

* From 0 to tested\_negative
* From 1 to tested\_positive

e.

* Class 0 = 500 instances
* Class 1 = 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.

=== Confusion Matrix ===

a b <-- classified as

420 80 | a = tested\_negative

111 157 | b = tested\_positive

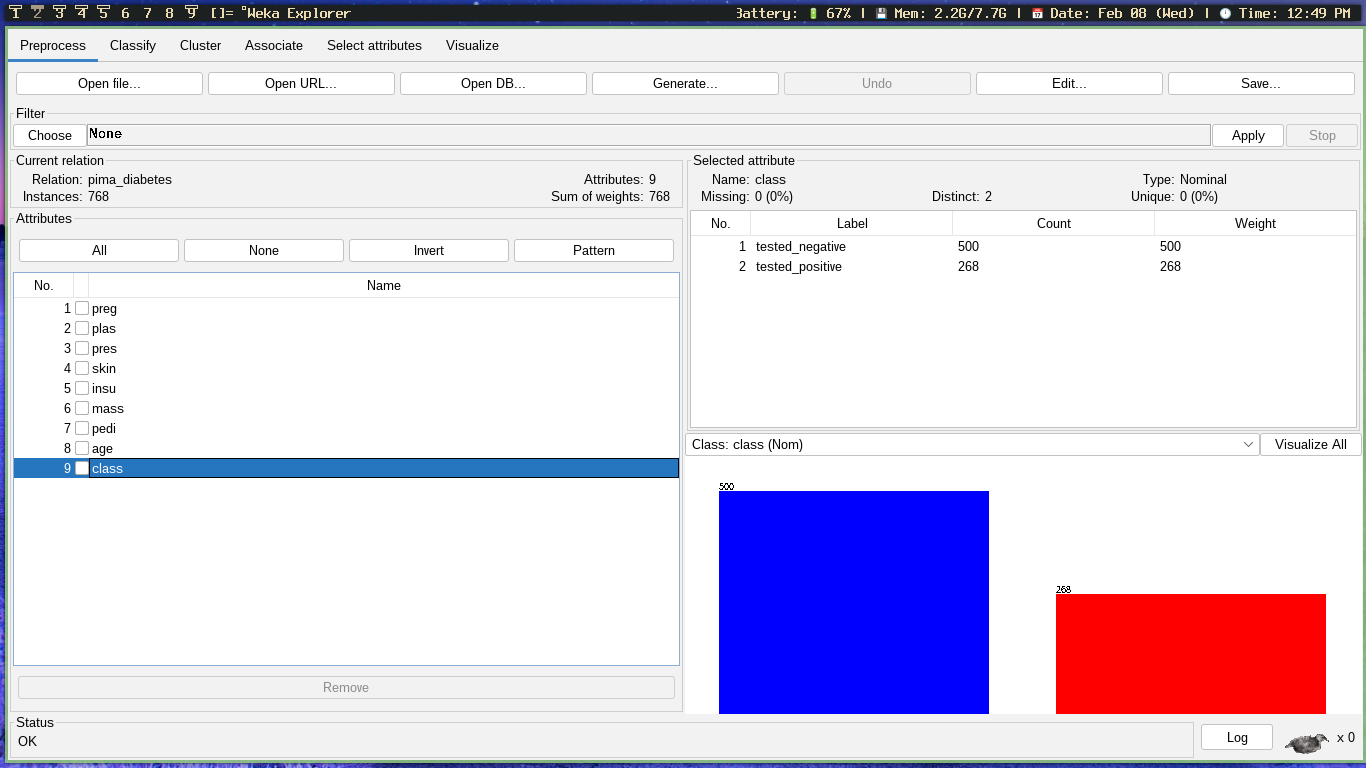
Details of random forest: Bagging with 100 iterations

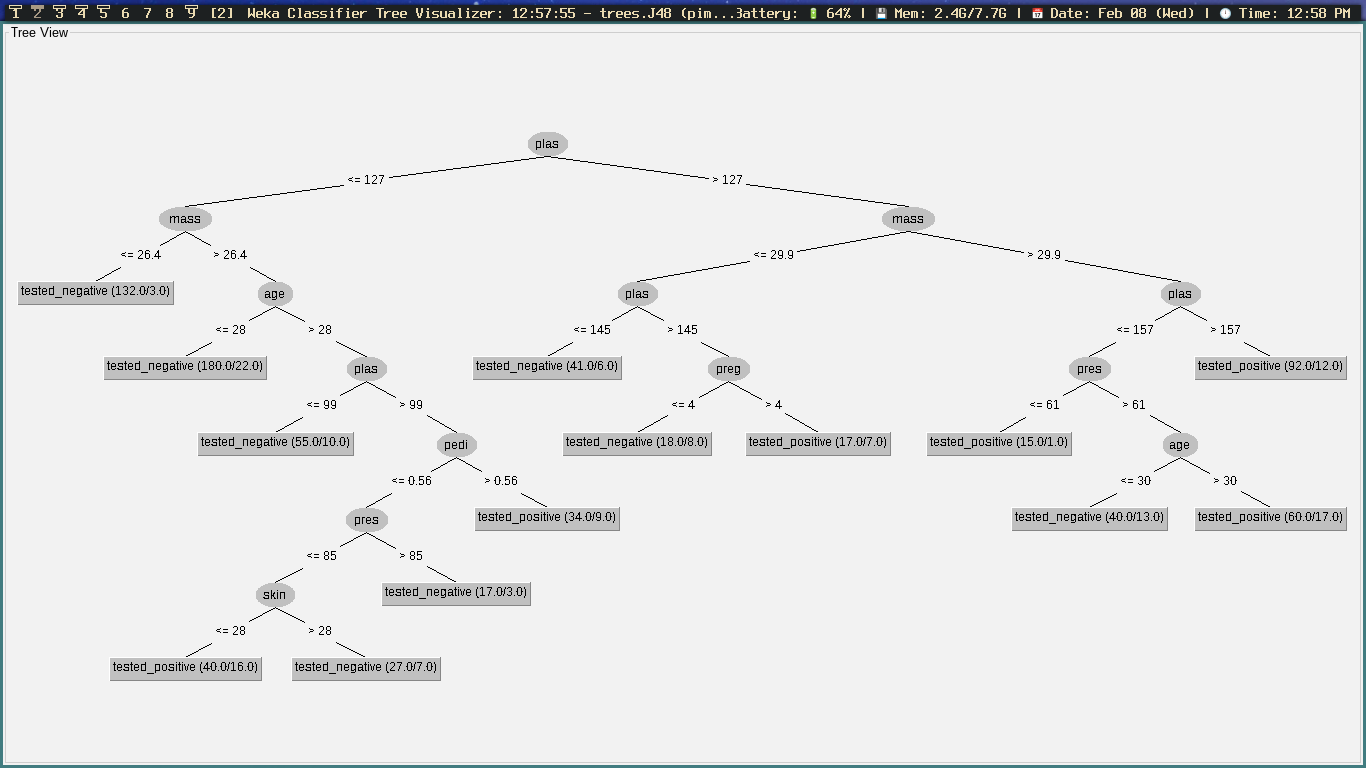
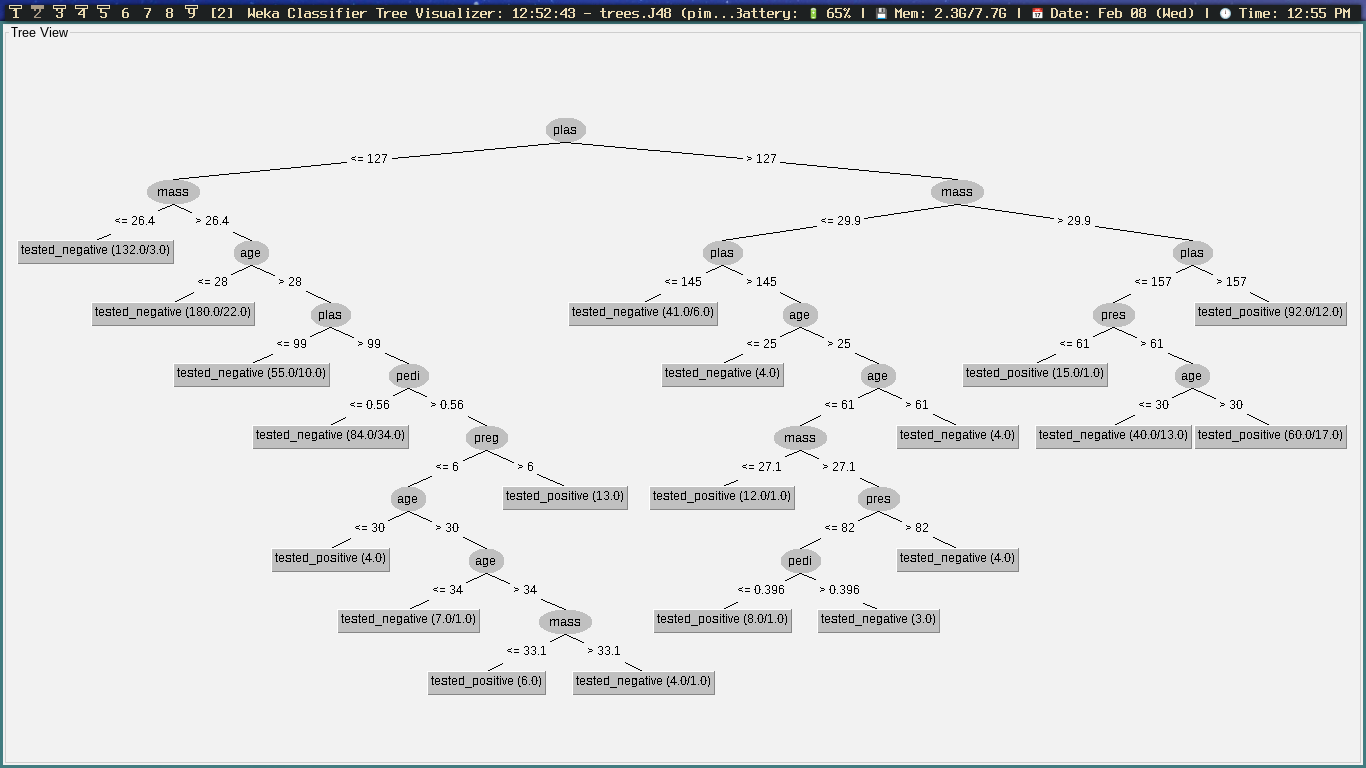
Time to build model: 0.5 seconds

Correctly classified instances: 577

**Screenshots with Question Number**

3.

  
6.

  
8.  
  
  
  
11.

