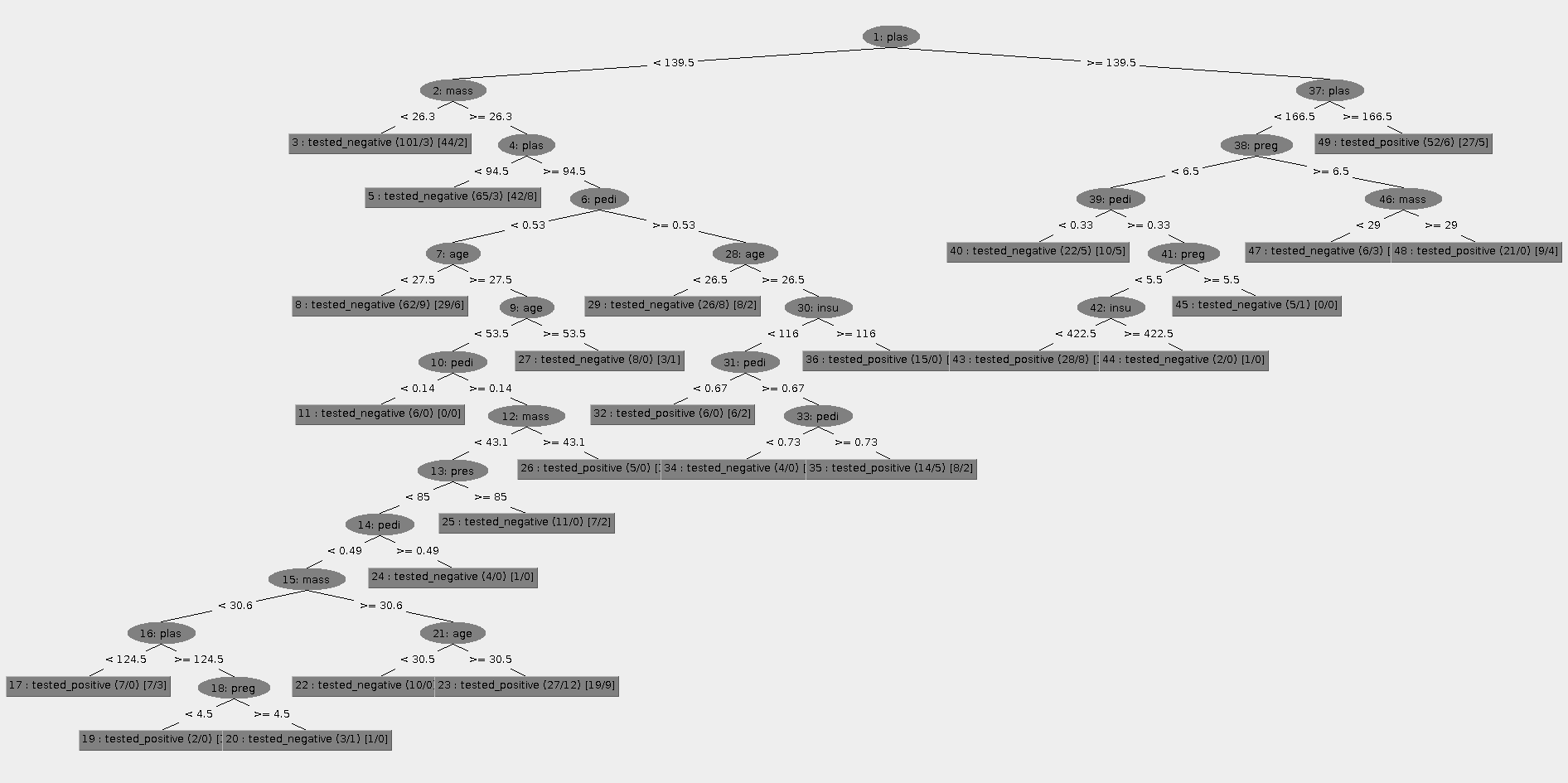
Considering the diabetes.arff dataset,

REP Tree with 10 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1

Relation: pima\_diabetes

Instances: 768

Attributes: 9

preg

plas

pres

skin

insu

mass

pedi

age

class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

REPTree

============

plas < 139.5

| mass < 26.3 : tested\_negative (101/3) [44/2]

| mass >= 26.3

| | plas < 94.5 : tested\_negative (65/3) [42/8]

| | plas >= 94.5

| | | pedi < 0.53

| | | | age < 27.5 : tested\_negative (62/9) [29/6]

| | | | age >= 27.5

| | | | | age < 53.5

| | | | | | pedi < 0.14 : tested\_negative (6/0) [0/0]

| | | | | | pedi >= 0.14

| | | | | | | mass < 43.1

| | | | | | | | pres < 85

| | | | | | | | | pedi < 0.49

| | | | | | | | | | mass < 30.6

| | | | | | | | | | | plas < 124.5 : tested\_positive (7/0) [7/3]

| | | | | | | | | | | plas >= 124.5

| | | | | | | | | | | | preg < 4.5 : tested\_positive (2/0) [1/0]

| | | | | | | | | | | | preg >= 4.5 : tested\_negative (3/1) [1/0]

| | | | | | | | | | mass >= 30.6

| | | | | | | | | | | age < 30.5 : tested\_negative (10/0) [3/0]

| | | | | | | | | | | age >= 30.5 : tested\_positive (27/12) [19/9]

| | | | | | | | | pedi >= 0.49 : tested\_negative (4/0) [1/0]

| | | | | | | | pres >= 85 : tested\_negative (11/0) [7/2]

| | | | | | | mass >= 43.1 : tested\_positive (5/0) [1/1]

| | | | | age >= 53.5 : tested\_negative (8/0) [3/1]

| | | pedi >= 0.53

| | | | age < 26.5 : tested\_negative (26/8) [8/2]

| | | | age >= 26.5

| | | | | insu < 116

| | | | | | pedi < 0.67 : tested\_positive (6/0) [6/2]

| | | | | | pedi >= 0.67

| | | | | | | pedi < 0.73 : tested\_negative (4/0) [3/0]

| | | | | | | pedi >= 0.73 : tested\_positive (14/5) [8/2]

| | | | | insu >= 116 : tested\_positive (15/0) [12/8]

plas >= 139.5

| plas < 166.5

| | preg < 6.5

| | | pedi < 0.33 : tested\_negative (22/5) [10/5]

| | | pedi >= 0.33

| | | | preg < 5.5

| | | | | insu < 422.5 : tested\_positive (28/8) [13/6]

| | | | | insu >= 422.5 : tested\_negative (2/0) [1/0]

| | | | preg >= 5.5 : tested\_negative (5/1) [0/0]

| | preg >= 6.5

| | | mass < 29 : tested\_negative (6/3) [1/0]

| | | mass >= 29 : tested\_positive (21/0) [9/4]

| plas >= 166.5 : tested\_positive (52/6) [27/5]

Size of the tree : 49

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 578 75.2604 %

Incorrectly Classified Instances 190 24.7396 %

Kappa statistic 0.438

Mean absolute error 0.3272

Root mean squared error 0.4289

Relative absolute error 71.9842 %

Root relative squared error 89.9782 %

Total Number of Instances 768

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.846 0.422 0.789 0.846 0.817 0.766 tested\_negative

0.578 0.154 0.668 0.578 0.62 0.766 tested\_positive

Weighted Avg. 0.753 0.328 0.747 0.753 0.748 0.766

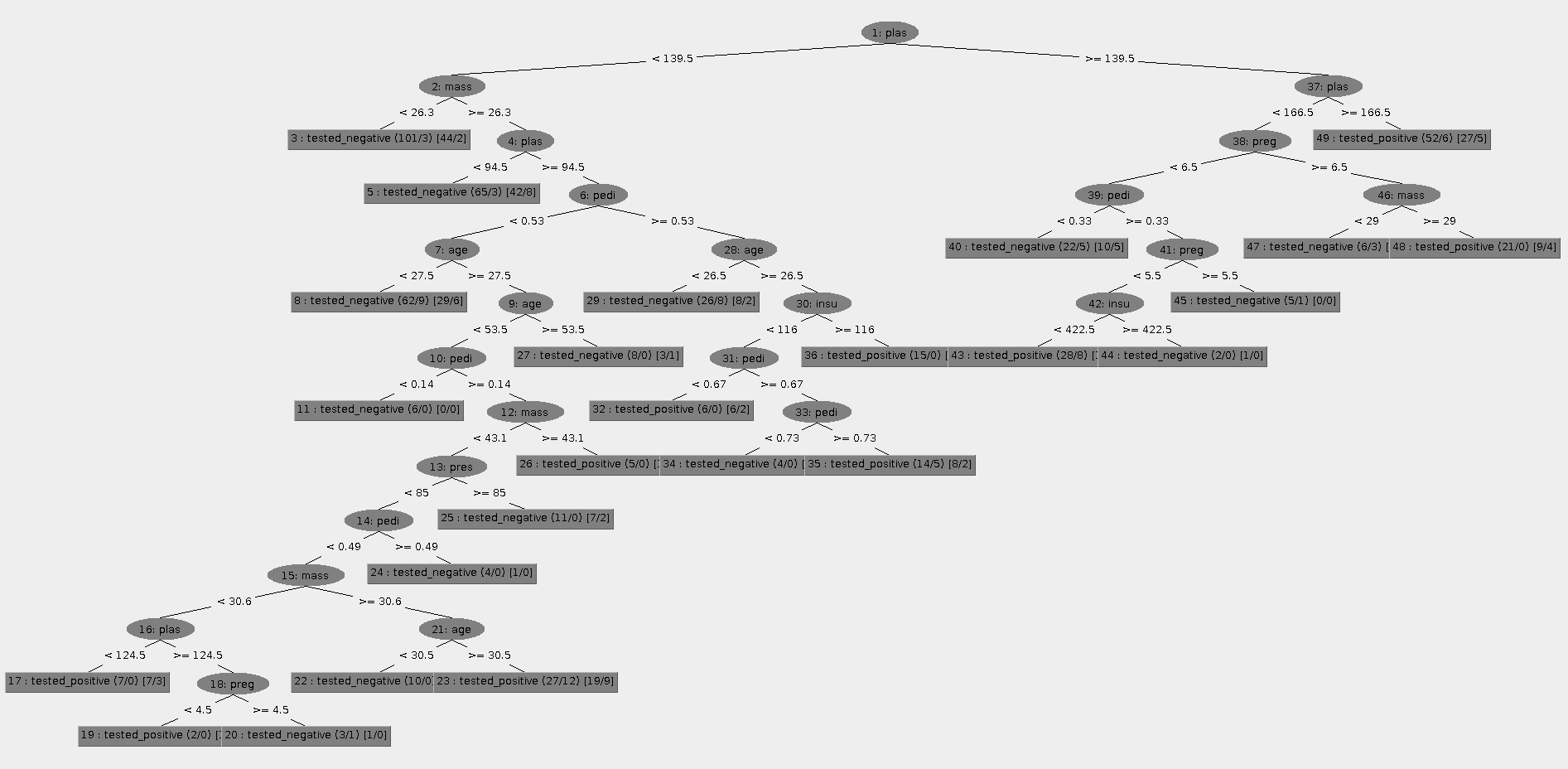
=== Confusion Matrix ===

a b <-- classified as

423 77 | a = tested\_negative

113 155 | b = tested\_positive

REP Tree with 20 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1

Relation: pima\_diabetes

Instances: 768

Attributes: 9

preg

plas

pres

skin

insu

mass

pedi

age

class

Test mode:20-fold cross-validation

=== Classifier model (full training set) ===

REPTree

============

plas < 139.5

| mass < 26.3 : tested\_negative (101/3) [44/2]

| mass >= 26.3

| | plas < 94.5 : tested\_negative (65/3) [42/8]

| | plas >= 94.5

| | | pedi < 0.53

| | | | age < 27.5 : tested\_negative (62/9) [29/6]

| | | | age >= 27.5

| | | | | age < 53.5

| | | | | | pedi < 0.14 : tested\_negative (6/0) [0/0]

| | | | | | pedi >= 0.14

| | | | | | | mass < 43.1

| | | | | | | | pres < 85

| | | | | | | | | pedi < 0.49

| | | | | | | | | | mass < 30.6

| | | | | | | | | | | plas < 124.5 : tested\_positive (7/0) [7/3]

| | | | | | | | | | | plas >= 124.5

| | | | | | | | | | | | preg < 4.5 : tested\_positive (2/0) [1/0]

| | | | | | | | | | | | preg >= 4.5 : tested\_negative (3/1) [1/0]

| | | | | | | | | | mass >= 30.6

| | | | | | | | | | | age < 30.5 : tested\_negative (10/0) [3/0]

| | | | | | | | | | | age >= 30.5 : tested\_positive (27/12) [19/9]

| | | | | | | | | pedi >= 0.49 : tested\_negative (4/0) [1/0]

| | | | | | | | pres >= 85 : tested\_negative (11/0) [7/2]

| | | | | | | mass >= 43.1 : tested\_positive (5/0) [1/1]

| | | | | age >= 53.5 : tested\_negative (8/0) [3/1]

| | | pedi >= 0.53

| | | | age < 26.5 : tested\_negative (26/8) [8/2]

| | | | age >= 26.5

| | | | | insu < 116

| | | | | | pedi < 0.67 : tested\_positive (6/0) [6/2]

| | | | | | pedi >= 0.67

| | | | | | | pedi < 0.73 : tested\_negative (4/0) [3/0]

| | | | | | | pedi >= 0.73 : tested\_positive (14/5) [8/2]

| | | | | insu >= 116 : tested\_positive (15/0) [12/8]

plas >= 139.5

| plas < 166.5

| | preg < 6.5

| | | pedi < 0.33 : tested\_negative (22/5) [10/5]

| | | pedi >= 0.33

| | | | preg < 5.5

| | | | | insu < 422.5 : tested\_positive (28/8) [13/6]

| | | | | insu >= 422.5 : tested\_negative (2/0) [1/0]

| | | | preg >= 5.5 : tested\_negative (5/1) [0/0]

| | preg >= 6.5

| | | mass < 29 : tested\_negative (6/3) [1/0]

| | | mass >= 29 : tested\_positive (21/0) [9/4]

| plas >= 166.5 : tested\_positive (52/6) [27/5]

Size of the tree : 49

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 556 72.3958 %

Incorrectly Classified Instances 212 27.6042 %

Kappa statistic 0.3741

Mean absolute error 0.3336

Root mean squared error 0.4393

Relative absolute error 73.3897 %

Root relative squared error 92.1634 %

Total Number of Instances 768

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.822 0.459 0.77 0.822 0.795 0.757 tested\_negative

0.541 0.178 0.62 0.541 0.578 0.757 tested\_positive

Weighted Avg. 0.724 0.361 0.717 0.724 0.719 0.757

=== Confusion Matrix ===

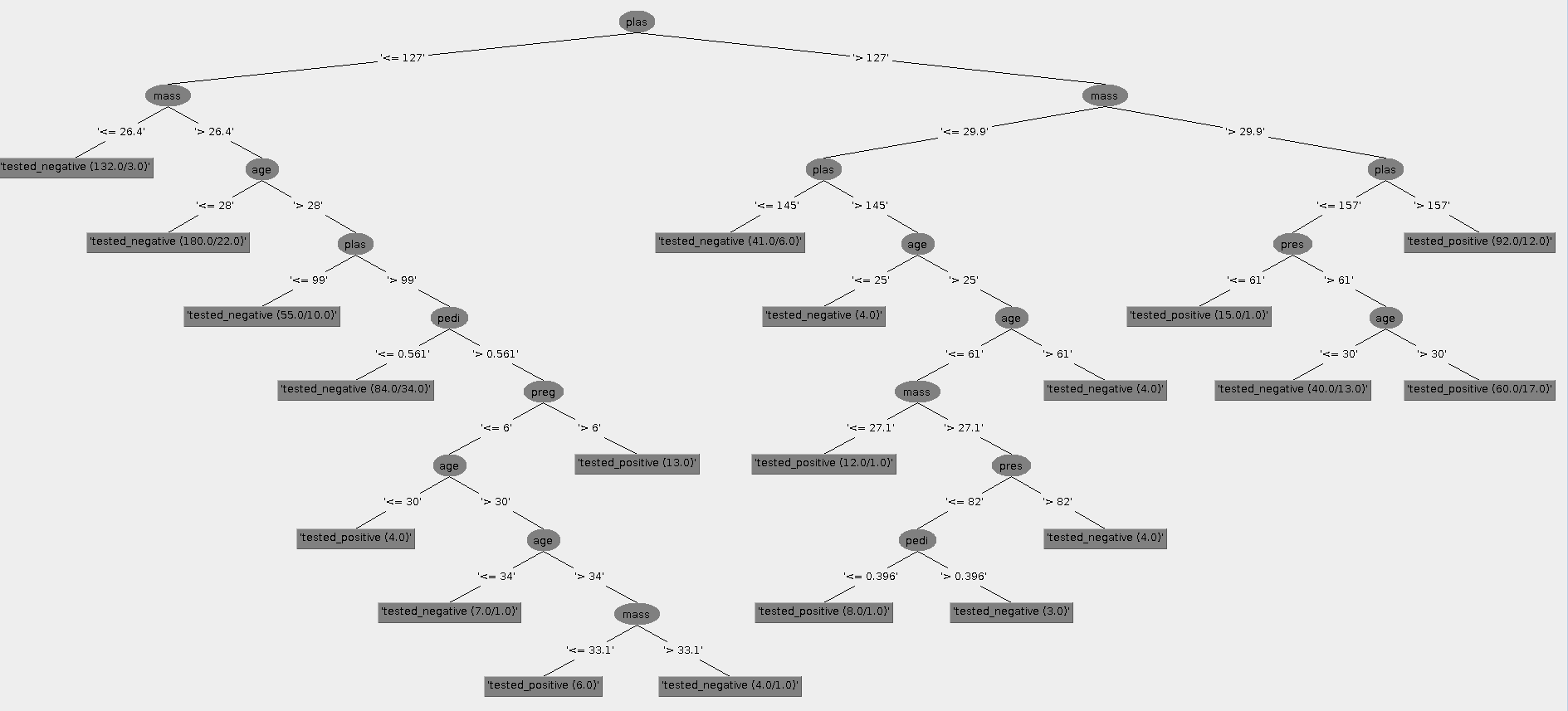
a b <-- classified as

411 89 | a = tested\_negative

123 145 | b = tested\_positive

**Both REP Trees produce the same decision tree as the algorithm is the same. The accuracy of classification is better in 10-fold REP tree (75.26%) compared to 20-fold REP tree (73.39%).**

J48 with 10 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: pima\_diabetes

Instances: 768

Attributes: 9

preg

plas

pres

skin

insu

mass

pedi

age

class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

------------------

plas <= 127

| mass <= 26.4: tested\_negative (132.0/3.0)

| mass > 26.4

| | age <= 28: tested\_negative (180.0/22.0)

| | age > 28

| | | plas <= 99: tested\_negative (55.0/10.0)

| | | plas > 99

| | | | pedi <= 0.561: tested\_negative (84.0/34.0)

| | | | pedi > 0.561

| | | | | preg <= 6

| | | | | | age <= 30: tested\_positive (4.0)

| | | | | | age > 30

| | | | | | | age <= 34: tested\_negative (7.0/1.0)

| | | | | | | age > 34

| | | | | | | | mass <= 33.1: tested\_positive (6.0)

| | | | | | | | mass > 33.1: tested\_negative (4.0/1.0)

| | | | | preg > 6: tested\_positive (13.0)

plas > 127

| mass <= 29.9

| | plas <= 145: tested\_negative (41.0/6.0)

| | plas > 145

| | | age <= 25: tested\_negative (4.0)

| | | age > 25

| | | | age <= 61

| | | | | mass <= 27.1: tested\_positive (12.0/1.0)

| | | | | mass > 27.1

| | | | | | pres <= 82

| | | | | | | pedi <= 0.396: tested\_positive (8.0/1.0)

| | | | | | | pedi > 0.396: tested\_negative (3.0)

| | | | | | pres > 82: tested\_negative (4.0)

| | | | age > 61: tested\_negative (4.0)

| mass > 29.9

| | plas <= 157

| | | pres <= 61: tested\_positive (15.0/1.0)

| | | pres > 61

| | | | age <= 30: tested\_negative (40.0/13.0)

| | | | age > 30: tested\_positive (60.0/17.0)

| | plas > 157: tested\_positive (92.0/12.0)

Number of Leaves: 20

Size of the tree: 39

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 567 73.8281 %

Incorrectly Classified Instances 201 26.1719 %

Kappa statistic 0.4164

Mean absolute error 0.3158

Root mean squared error 0.4463

Relative absolute error 69.4841 %

Root relative squared error 93.6293 %

Total Number of Instances 768

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.814 0.403 0.79 0.814 0.802 0.751 tested\_negative

0.597 0.186 0.632 0.597 0.614 0.751 tested\_positive

Weighted Avg. 0.738 0.327 0.735 0.738 0.736 0.751

=== Confusion Matrix ===

a b <-- classified as

407 93 | a = tested\_negative

108 160 | b = tested\_positive

**The J48 tree produces a much simpler tree compared to REP Tree but the accuracy is 73.83% which is lower than the REP Tree with same number of cross-validation folds. The accuracy of classification of REP Tree is usually expected to be higher than that of J48, as REP Tree uses reduced-error pruning in addition to the C4.5 algorithm used by J48 Tree.**